Distinguishing between Classical and Popular Music: The Influence of Instrument, Harmony, Voice Leading and More

Junyi Zhu*
junyizhu@andrew.cmu.edu

December 8, 2019

Abstract

We explored the effects of music-related factors on rating musical stimuli as classical or popular, using mixed effects hierarchical models on the data from Jimenez & Rossi (2012). To derive the models, both best subsets and automated methods are used to find the best fixed and random effects. We found that instruments have the strongest effect on ratings: while string quartets and Harmony I-V-VI are associated with higher classical ratings, guitar is associated with higher popular ratings. In addition, various aspects in musical knowledge and self-identification as a musician also have significant impact on both ratings. Overall, popular ratings are mainly affected by instrument while classical ratings are influenced by a combination of factors in a more complicated manner.

1 Introduction

Whereas there have been extensive studies on distinguishing classical and popular music listeners (Prieto-Rodríguez and Fernández-Blanco, 2000) or on the effects of pitch and rhythmic information on musical identification (Halpern and Bartlett, 2010), the effects of instrument, harmony and voice leading on such musical identification have seldom been studied. The dataset from Jimenez & Rossi (2012) is collected in a designed experiment and is intended to measure the influence of instrument, harmonic motion and voice leading on listeners' identification of music as "classical" or "popular". In the dataset, 70 participants each rated 36 musical stimuli on both classical scale and popular scale of 1 to 10, and music-related characteristics of the participants are recorded.

Utilizing this dataset, this paper presents the best hierarchical mixed effects models that predict ratings of music pieces as "classical" and as "popular" and investigates which factors most influence the ratings. In particular, we are interested in addressing the main research question by answering the following questions:

- Out of the three experimental factors - Instrument, Harmony and Voice Leading - which one or which combination has the strongest influence on musical ratings?
- Are there differences in the additional covariates that drive classical vs. popular ratings?
- Are there differences in the way that musicians and non-musicians identify classical music?

*MSP, Department of Statistics & Data Science, Carnegie Mellon University

2 Methods

As the first step, we examine the original dataset and address various problems such as missing values, miscoding and skewed variables. Table 1 below contains the definitions of all variables in the original dataset, and the starred variables were taken out of the dataset after the data cleaning process. Readers should refer to Jimenez & Rossi (2012) for detailed definitions, criteria of selected musical stimuli, information about participants, and so forth of the raw data.

Table 1: Variable definitions of Data from Jimenez & Rossi (2012)

Variable Name	Description
Classical	On a scale of 1 to 10, how classical does the music sound?
Popular	On a scale of 1 to 10, how popular does the music sound?
Subject	Unique Subject ID
Harmony	Harmonic Motion: I-IV-V, I-V-IV, I-V-VI, and IV-I-V
Instrument	Instrument: guitar, piano, string quartets
Voice	Voice Leading: contrary motion, parallel 3rd, parallel 5th
Selfdeclare	Are you a musician? (1-6, 1=not at all)
OMSI	Score on a test of musical knowledge
X16.minus.17	Measure of listener's ability to distinguish classical vs popular music
ConsInstr	How much did you concentrate on the instrument? (0-5, 0=not at all)
ConsNotes	How much did you concentrate on the notes? (0-5, 0=not at all)
Instr.minus.Notes*	Difference between the previous two variables
PachListen	How familiar are you with Pachelbel's Canon in D (0-5, 0=not at all)
ClsListen	How much do you listen to classical music? (0-5, 0=not at all)
KnowRob	Have you heard of Rob Paravonian's Pachelbel Rant? (0-5, 0=not at all)
KnowAxis	Have you heard of Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music? (0-5, 0=not at all)
X1990s2000s	How much do you listen to pop and rock from the 90's and 2000's? (0-5, 0=not at all)
X.minus	Difference between `X1990s2000s` and a similar variable referring to 60's and 70's pop and rock.
CollegeMusic	Have you taken music classes in college (0=no, 1=yes)
APTheory	Did you take AP Music Theory class in High School (0=no, 1=yes)
Composing	Have you done any music composing (0-5, 0=not at all)
PianoPlay	Do you play piano (0-5, 0=not at all)
GuitarPlay	Do you play guitar (0-5, 0=not at all)
X1stInstr*	How proficient are you at your first musical instrument (0-5, 0=not at all)
X2ndInstr*	Same, for second musical instrument (0-5, 0=not at all)
first12*	Which instrument was presented to the subject in the first 12 stimuli?

For missing values, we first removed the problematic or unnecessary columns (starred in Table 1 above), and then removed the rows with missing values in `Classical` or `Popular`. In addition, median imputation is performed for the rest of the missing values, because simply dropping the rows with any missing values would leave us to only about 60% of the original dataset. Note that median was preferred over mean, because most columns are categorical variables with integer-leveled factors. Throughout the analysis, we should keep in mind that single imputation would alter the variable distribution and lower the variance of a variable, and thus understate the uncertainty of a variable. Next, for miscoding, we inspected all variables and ascertained that all variables meet the specification (e.g. number of levels, maximum and minimum constraints, etc.) listed in the data description.

After data cleaning, exploratory data analysis is performed by looking at the distribution of, correlation among and boxplots (grouped by Classical and Popular ratings) of all remaining variables[1]. Variable transformation is performed when necessary. With the adequate knowledge about the variables at hand, we move onto the model building procedure for predictions of musical ratings.

First of all, in Section 3.1, we examined the influence of the three main experimental factors (Instrument, Harmony and Voice Leading) on musical ratings. Using classical and popular ratings as response variables respectively, we built two linear models with the three factors and all possible interactions between them as predictors, conducted variable selection by comparing AIC/BIC values, forward selection and backward elimination, and decided which interactions should be selected as fixed effects. All three experimental factors are included in the model for comparing their effect sizes. Thereafter, random effects were selected by both comparing all possible subsets and utilizing automated method of back-fitting fixed effects and forward-fitting random effects, allowing the possibility of including random intercepts and random slopes. In particular, to contrast the two best hierarchical mixed effects models, we compared the effect sizes among predictors and among levels within predictors to evaluate the three secondary hypotheses that (1) Instrument exerts the strongest influence among the three design factors and that (2) Harmony I-V-VI or (3) Voice of contrary motion is strongly associated with classical ratings.

Next, in Section 3.2, we determined which additional individual covariates should be added to both models as fixed effects respectively, based on the best models developed in the previous section. The same automated step-wise method from Section 3.1 was used to develop the best linear fixed effects models. Once the fixed effects were selected, we again employed the same automated method as in Section 3.1 to select random effects, allowing the model to be hierarchical. After using all possible covariates to build the best hierarchical mixed effects models for both classical and popular ratings, we compared and interpreted the differences in the significant predictors that drive classical vs. popular ratings.

Last, in Section 3.3, we investigated the hypothesis that people who self-identify as musicians may be influenced by factors that do not influence non-musicians. Specifically, we dichotomized `Selfdeclare`, a categorical variable of 6 ordered numeric levels, so that about half the participants are categorized as self-declared musicians (values 1) and half not (values 0). After the dichotomy, we selected fixed effects with automated step-wise method on all possible two-way interactions between the musician variable and other predictors selected for the classical model. We used both AIC and BIC values and eventually selected the models by BIC after comparison. As we are interested only in the effect of self-identification as a musician, no random effects were considered. Also, the above mentioned procedure was dichotomized twice, between different levels in `Selfdeclare` (level 2 and 3, level 3 and 4), to investigate whether the results are sensitive to where we dichotomize.

For all models, ANOVA tests and exact log likelihood ratio tests were performed when necessary to select nested maximum likelihood models (ML) and restricted maximum likelihood (REML) models respectively. Throughout the model building process, we strove to seek a balance between rigorous statistical correctness and sensible musical interpretations. All above mentioned analyses were carried out by the R language and environment for statistical computing (R core team, 2017).

[1]Selected graphs will be presented in the Results Section. Please refer to Appendix B for related R codes.

3 Results

To clean the original dataset, we first removed the column `Instr.minus.Notes` due to perfect collinearity with `ConsInstr` and `ConsNotes`[2], the column `first12` due to unnecessity, and the columns `X1stInstr` and `X2ndInstr` due to significant number of missing values[3]. Thereafter, we removed 27 rows with missing values in `Classical` or `Popular`, since the values in these two columns are needed for response variables for the modelling process. Finally, we performed median imputation for missing values in the columns `ConsNotes`, `PachListen`, `ClsListen`, `KnowRob`, `KnowAxis`, `X1990s2000s`, `X.minus`, `CollegeMusic`, `NoClass`, `APTheory` and `Composing`.

For miscoding, `ConsInstr` was found to contain non-integer values that make it a 14-level factor instead of a 5-level factor as the dataset claimed. Therefore, all non-integer values for `ConsInstr` are rounded to integer values.

Next, based on the histograms of the variables, a logarithm transformation is performed to `OMSI` in order to force the covariate into a roughly normal distribution. As shown in Figure 1 below, `OMSI` is strongly skewed to the right, and looks reasonably normal after the transformation.

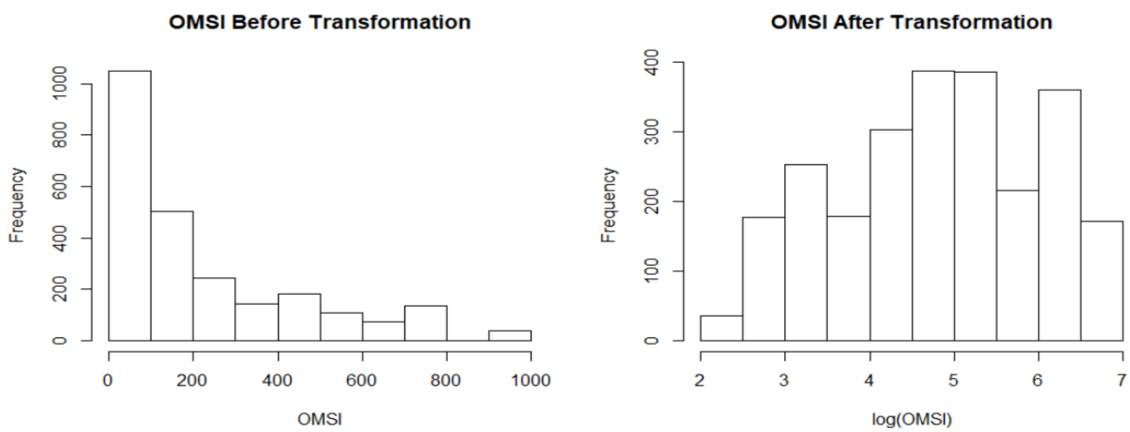

Figure 1: Transformation of the OMSI Variable

Furthermore, the correlation table between the continuous variables and the categorical variables with more than 2 numerical levels is inspected is presented in Figure 2, where a deeper orange indicates a more positive correlation and a deeper blue indicates a more negative correlation. We obtain a few observations listed below that more or less align with common sense:

- Being more active in composing, playing guitar and playing piano are all moderately positively correlated with self-declaring more as a musician (0.58, 0.48, 0.61 respectively) and with higher score on OMSI, the test score of musical knowledge (0.54, 0.44, 0.68 respectively).

- Self-declaring more as a musician is highly positively correlated with having a high OMSI score (0.77).

- Being more active in playing guitar (but not piano or string) is highly positively correlated with having a high score on Composing (0.68).

[2]See variable description in Table 1.
[3]These two columns contain 1,512 and 2,196 missing values respectively, from a total of 2520 rows.

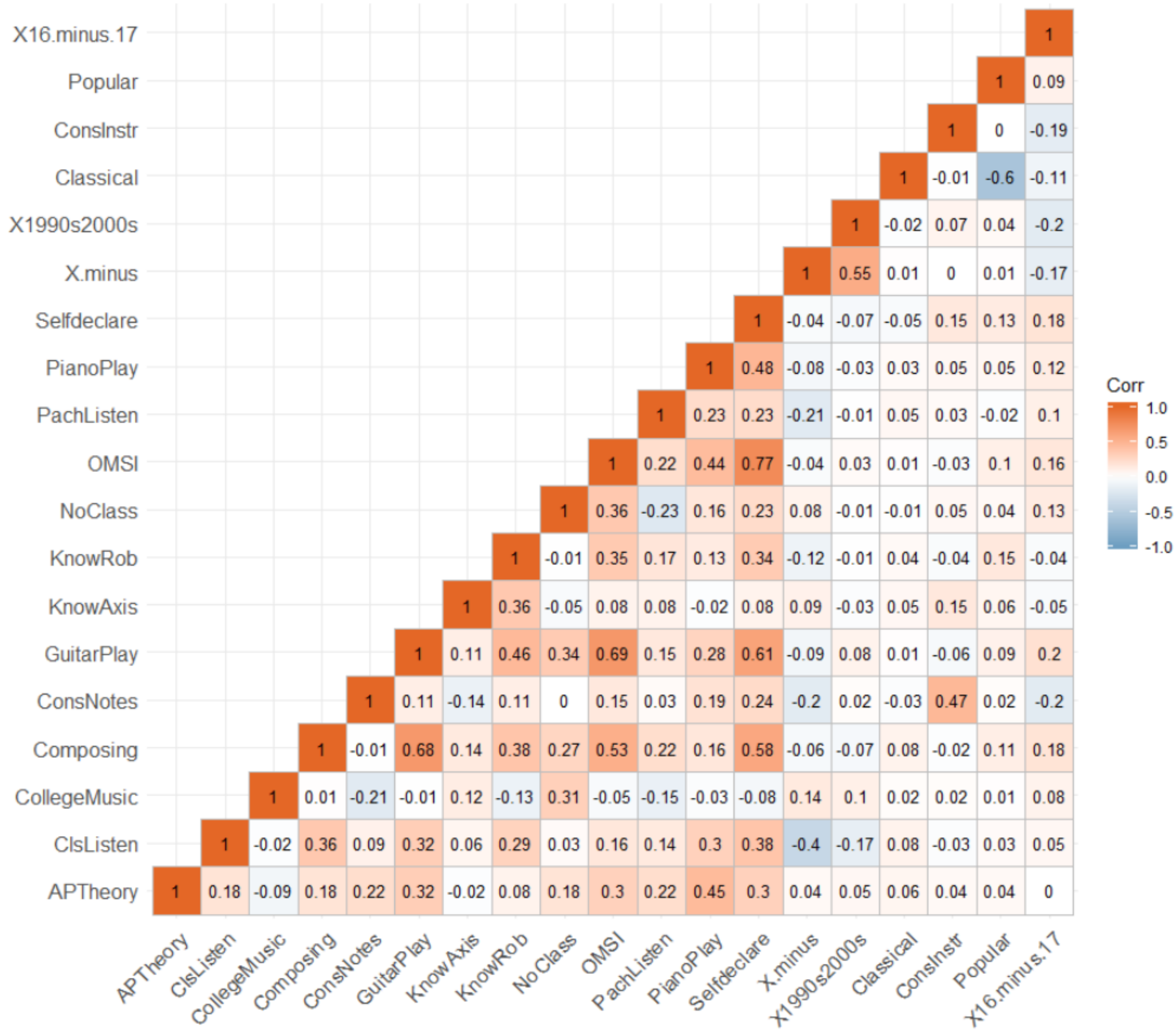

Figure 2: Correlation Table

These observations are crucial in the model building process, as high correlation among the predictors should be noted to avoid the problem of collinearity as much as possible. In other words, when the model includes two predictors that are known to be highly correlated, one of them should be dropped so that the underlying model assumption is satisfied.

Apart from the correlation among the predictors, the negative correlation (-0.60) between `Classical` and `Popular` should also be noted. In other words, music pieces that are rated higher on the classical scale have a relatively lower score on the popular rating scale. To some extent, this particular observation suggests that when a predictor is positively associated with higher classical ratings, it is likely to be positively associated with lower popular ratings at the same time. As we will see in the following sections, this observation holds true for quite a few of the predictors, especially for the three main experimental factors: Instrument, Harmony, and Voice Leading.

3.1 The Influence of Instrument, Harmony and Voice Leading

In this section, we investigate what experimental factor, or combination of factors, has the strongest influence on musical ratings. In particular, we present and interpret the two hierarchical mixed effects models that best predict classical ratings and popular ratings respectively[4].

3.1.1 Predicting Classical Ratings Using the Three Main Effects

For classical ratings, the fixed effects Instrument, Harmony, Voice and the interaction between Harmony and Voice are selected. For the random effects, the random intercept that varies by subject and the random slopes for Instrument and Harmony that also vary by subject are included. The model assumes that there is correlation between the random effects of Instrument by Subject and the random effects of Harmony by Subject.

Selected part of the classical model summary is printed out below:

```
Formula: Classical ~ 1 + Instrument + Harmony + Voice + Harmony * Voice +
    (1 | Subject) + (Instrument + Harmony | Subject)

Random effects:
 Groups     Name             Variance  Std.Dev.  Corr
 Subject    (Intercept)      7.924e-08 0.0002815
 Subject.1  (Intercept)      2.544e+00 1.5951128
            Instrumentpiano  1.632e+00 1.2776737 -0.39
            Instrumentstring 3.510e+00 1.8735032 -0.57  0.66
            HarmonyI-V-IV    4.084e-02 0.2020911  0.70 -0.67 -0.44
            HarmonyI-V-VI    1.588e+00 1.2602350 -0.05 -0.27 -0.42  0.21
            HarmonyIV-I-V    9.000e-03 0.0948693  0.18 -0.38  0.14  0.27  0.14
 Residual                    2.418e+00 1.5549828
Number of obs: 2493, groups:  Subject, 70

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                4.25226    0.22329  19.044
Instrumentpiano            1.37008    0.17102   8.011
Instrumentstring           3.12747    0.23652  13.223
HarmonyI-V-IV              0.15517    0.15465   1.003
HarmonyI-V-VI              1.13872    0.21443   5.310
HarmonyIV-I-V             -0.13338    0.15254  -0.874
Voicepar3rd               -0.27042    0.15217  -1.777
Voicepar5th               -0.23646    0.15253  -1.550
HarmonyI-V-IV:Voicepar3rd -0.36489    0.21566  -1.692
HarmonyI-V-VI:Voicepar3rd -0.68009    0.21590  -3.150
HarmonyIV-I-V:Voicepar3rd  0.48537    0.21557   2.252
HarmonyI-V-IV:Voicepar5th -0.18919    0.21609  -0.876
HarmonyI-V-VI:Voicepar5th -0.42576    0.21584  -1.973
HarmonyIV-I-V:Voicepar5th  0.07525    0.21545   0.349
```

[4]Refer to Appendix C for codes for model selection process and residual plots.

Using the outputs from above, the fixed effects for the classical model are interpreted below, while keeping all other variables constant. Since the response variable is a categorical variable, the interpretations is based on comparisons with regards to the baseline (i.e. the intercept).

- Baseline: the music played to the subject uses guitar as the instrument, followed Harmony I-IV-V, and the singer's voice is categorized as Contrary, with coefficient being 4.25226 and the standard error (denoted as s.e. below) being 0.22329.

- `Instrument`: when compared to the base level of having guitar as the instrument in the music piece, having piano is associated with a 1.370 *increase* in the classical rating (s.e.=0.171) whereas having string quartets is associated with a 3.127 *increase* in the classical rating (s.e.=0.237).

- `Harmony`: when compared to the base level of having Harmony I-IV-V as the harmonic motion in the music piece: having Harmony I-V-IV is associated with a 0.155 *increase* in the classical rating (s.e.=0.155); having Harmony I-V-VI is associated with a 1.139 *increase* in the classical rating (s.e.=0.214), having Harmony I-V-IV is associated with a 0.133 *decrease* in the classical rating (s.e.=0.152). Note that confidence intervals for all three coefficients contain zero, meaning that these interpretations are not statistically significant.

- `Voice`: when compared to the base level of having the Voice Contrary in the music piece: having Voice Parallel 3rd is associated with a 0.270 *decrease* in the classical rating (s.e.=0.152), whereas having Voice Parallel 5th is associated with a 0.236 *decrease* in the classical rating (s.e.=0.153).

- To interpret the interaction terms, `HarmonyI-V-IV:Voicepar3rd` is used as an example. Comparing with `HarmonyI-IV-V` (base level) and fixing `Voice` to be Parallel 3rd, the music being `HarmonyI-V-IV` is associated with a 0.635 *decrease* in the classical rating (coefficient=(-0.36489)+(-0.27042)=-0.63531). Comparing with `Voice-contrary` (base level) and fixing `Harmony` to be I-IV-V, the voice being Parallel 3rd is associated with a 0.20972 *decrease* in the classical rating (coef=(-0.36489)+(0.15517)=-0.20972).

Next, the random effects included in the classical model are interpreted below. In particular, we look at the sizes of the variances of all random effects in the model, with respect to each other and with respect to the estimated residual variance.

- With respect to each other, the variance of the random effects for Harmony is in general smaller than the variance of the random effects for Instruments, meaning that when in comparison, the random effects of Instruments vary more by subjects than the random effects of Harmony.

- With respect to the estimated residual variance, we see that the variance of the random effects for Instrument String (3.510) is bigger than the variance of the residual random effects (2.418), thus the comparison gives evidence that including the random effects for Instrument is valid. On the other hand, the variance of the random effects for all levels of Harmony (4.084e-02, 1.588, 9.000e-03) is smaller than the variance of the residual random effects (2.418), thus the random effects for Harmony might not be needed.

3.1.2 Predicting Popular Ratings Using the Three Main Effects

Interestingly, for popular ratings, the selected predictors are almost identical to the ones in the classical model. The fixed effects are Instrument, Harmony and Voice. The random effects, all varying by Subject, are the random intercept and the random slopes of Instrument and Harmony. Selected model summary is printed below for interpretation.

```
Formula: Popular ~ 1 + Instrument + Harmony + Voice + (1 | Subject) +
    (Instrument + Harmony | Subject)
Random effects:
 Groups    Name             Variance Std.Dev. Corr
 Subject   (Intercept)      0.05542  0.2354
 Subject.1 (Intercept)      1.55805  1.2482
           Instrumentpiano  1.41997  1.1916   -0.20
           Instrumentstring 3.35051  1.8304   -0.33  0.72
           HarmonyI-V-IV    0.12412  0.3523    0.50 -0.14 -0.29
           HarmonyI-V-VI    0.91002  0.9539   -0.17 -0.21 -0.23 -0.39
           HarmonyIV-I-V    0.25885  0.5088   -0.28 -0.17 -0.09 -0.69 -0.16
 Residual                   2.49821  1.5806
Number of obs: 2493, groups:  Subject, 70
Fixed effects:
                 Estimate Std. Error t value
(Intercept)       6.57971    0.17620  37.342
Instrumentpiano  -0.94738    0.16249  -5.831
Instrumentstring -2.60529    0.23206 -11.227
HarmonyI-V-IV    -0.02677    0.09897  -0.270
HarmonyI-V-VI    -0.27371    0.14501  -1.888
HarmonyIV-I-V    -0.18568    0.10818  -1.716
Voicepar3rd       0.16544    0.07759   2.132
Voicepar5th       0.16199    0.07754   2.089
```

Similar to we did for the classical model, we interpret the fixed effects for the popular model by comparing the coefficients with the baseline while keeping all other variables constant.

- Baseline: the baseline is a mixture of using guitar as the instrument, following Harmony I-IV-V, and having a contrary voice leading. The coefficient is 6.57971 (s.e.=0.17620).
- `Instrument`: when compared to guitar, having piano in the music is associated with a 0.947 *decrease* in the popular ratings (s.e.=0.162), while having string quartets is associated with a 2.61 *decrease* in the popular ratings (s.e.=0.232).
- `Harmony`: when compared to Harmony I-IV-V: having I-V-IV is associated with a 0.02677 *decrease* in the popular ratings (s.e.=0.09897); having I-V-VI is associated with a 0.274 *decrease* in the popular ratings (s.e.=0.145); and having I-V-IV is associated with a 0.186 *decrease* in the popular ratings (s.e.=0.108). Again, none of the above is statistically significant.
- `Voice`: when compared to the base level of Voice Contrary, having Voice Parallel 3rd is associated with a 0.165 *increase* in the Popular rating (s.e.=0.0776), whereas having Voice Parallel 5th is associated with a 0.162 *increase* in the Popular rating (s.e.=0.0775).

As for the random effects in the popular model, we compare again the variances of random effects and the variance for the estimated residuals. The results are identical to those for the classical model: the random effects of Instruments vary more by subjects than the random effects of Harmony, including the random effects for Instrument is valid, and the random effects for Harmony might not be needed.

Why is Harmony not significant for the popular model? For both mixed effects models, the automated method used for random effects selection actually found that including Harmony (and thus the interaction of Harmony and Voice, or the random effect of Harmony) as a main effect is unnecessary. In other words, the automated method simply confirms that having the random intercept and allowing the random effect of Instrument to vary by subject is reasonable. However, since the research itself wants to explore the effects of Instrument, Harmony and Voice have on the ratings, we decided to keep Harmony and its random effects in the final model.

We also take note of the obvious opposite signs of the same predictors in the classical and the popular model. For example, while the string quartets level in instrument is strongly and statistically significantly associated with higher classical ratings, this level is statistically significantly associated with lower popular ratings. This result holds true for most levels in the main three predictors, even though the response variable is not binary. In other words, a subject can technically rate a music piece as high on both classical ratings and popular ratings. Yet this pattern of opposite signs suggests that the subjects implicitly assume that popular and classical are on two ends of a spectrum when they evaluate the musical stimuli. This pattern will be observed in the following subsections, in which we discuss the influence of each design factor.

Finally, we checked the appropriateness for our models. For both models, the appropriateness of the prediction model is checked by looking at 4 distinct residual plots: (1) the binned residual plots, (2) the plot of marginal residuals vs. fitted marginal values, (3) the normal QQ plot for conditional residuals and (4) the normal QQ plot for random effect residuals. Since the selected predictors for classical ratings and popular ratings are quite similar, the behaviors of all four types of residuals for both models look, unsurprisingly, almost identical. For both the classical and the popular model, there are about 5 outliers in the binned plot and one potential outlier on the right hand side in the three other plots. The marginal residuals are quite evenly distributed, and both conditional and random effect residuals follow the normal pattern pretty well, with slightly long tails on either side. Despite minor imperfections, overall, both of the chosen models seem to satisfy the underlying model assumptions to a reasonable degree, and thus are adequate in predicting the classical ratings and the popular ratings respectively.

3.1.3 Instrument: the Most Significant Predictor for Both Musical Ratings

Instrument exerts the greatest influence among all main experiment factors and their interactions. As shown in the previous two sections, the t-values for all levels in Instrument for both models are either greater than 2 or smaller than -2, indicating that the coefficients significantly differ from 0. Furthermore, the coefficients for levels in instrument are also largest when compared to the levels in other predictors. In summary, Instrument is not only significant in predicting both classical or popular music, but also it has a large impact on the musical ratings.

In particular, the classical ratings are highest when the instrument used in the music is a string quartet, followed by piano, and lastly guitar. On the contrary, the popular ratings are highest when the instrument used in the music is guitar, followed by piano, and lastly string - the complete opposite pattern when compared to the results of classical ratings. Additionally, instruments is the only significant predictor of popular ratings. Intuitively, the result aligns with our common sense, as piano and string quartets are typical components in classical music, and guitar has been widely used in popular music. This result also confirms one of the researchers' main hypotheses in that instrument indeed has the largest influence on the ratings of the musical stimuli.

Last but not least, it is worth mentioning that significant difference in the influence of instrument is found in the subjects. The results from both model suggests that allowing the random effect of Instrument to vary by subject is helpful in improving the model. Given the fact that the subjects have varying levels of musical expertise, it aligns with common sense that some subjects might be more influenced by the instrument, while others might spare more attention to other important features in the music.

3.1.4 The Special Effect of Harmonic Motion I-V-VI on Classical Ratings

Compared to Instrument, the harmonic motions have a weaker but more complicated influence on only classical ratings[5]. In particular, we want to investigate the harmonic progression I-V-VI, as it is the only significant level in Harmony that predicts classical ratings.

First of all, the harmonic progression I-V-VI in the music is associated with higher classical ratings (statistically significant, coef=1.14, t-value=5.31) but lower popular ratings (not statistically significant, t-value=-1.89). However, we found that the interaction between harmonic progression I-V-VI and parallel 3rd voice leading is actually significantly associated with lower classical ratings (coef=-0.68, t-value=2.25), while the interaction between harmonic progression IV-I-V and parallel 3rd voice leading is weakly associated with higher classical ratings (coef=0.49, t-value=-3.15). Again, this result aligns with researchers' hypothesis that the harmonic progression I-V-VI might be frequently rated as classical, because it is the beginning progression for Pachelbel's Canon in D, which many people have heard. To further investigate whether familiarity with one or the other (or both) of the Pachelbel rants/comedy bits matter in classical ratings, the researchers may want to consider an interaction between the `Harmony` and `PachListen`.

In addition, addressing the significant interaction terms between harmony and voice, the researchers may want to further look into why these two specific interactions between voice and harmonic motion would cause such an influence on classical ratings.

3.1.5 Three Levels in Voice Leading: Similar Effects on Classical Ratings

Referring back to both models, the effects of Voice Leading in parallel motions, whether in 3rd or in 5th, do not significantly differ from the effect of Voice Leading in contrary motions. More specifically speaking, the model produced by the step-wise method using both backward elimination and forward selection actually excluded Voice as a predictor, but Voice was manually added back into the model because it is one of the three design factors in the experiment.

Although none of the levels of Voice Leading are significant by themselves in predicting either classical or popular ratings, the interactions between Parallel 3rd in Voice Leading and I-V-VI or IV-I-V of harmonic motion are significant in predicting classical ratings. In addition, considering the baseline, when the musical stimulus is played in guitar with harmonic motion I-IV-V and voice leading of contrary motion, the model predicts the greatest significant increase in classical ratings.

In summary, the three levels of Voice Leading does not seem to differ in their effects on classical ratings, as none of them are significant in the prediction model.

[5] None of the levels in harmonic motion statistically affect popular ratings.

3.2 Additional Predictors for Both Classical and Popular Ratings

Building upon the best models from the previous section, we now use all possible predictors from the cleaned dataset to develop models that best predict musical ratings. We will briefly talk about the best hierarchical mixed effects models for both classical and popular ratings, and then compare and interpret the differences in the significant predictors that drive classical vs. popular ratings.

When we consider adding other individual covariates into the model, we found that while adding some variables improved both models that predict classical and popular ratings, adding other variables only improved the prediction of either classical or popular model. In addition, we note that adding the additional covariates does not change the interpretations of the three main effects. Due to the large amount of covariates included in both models, the full models will not be presented[6]. In this section, we will present a brief summary of which fixed effects are chosen to improve the prediction of musical ratings, given in Figure 3 below.

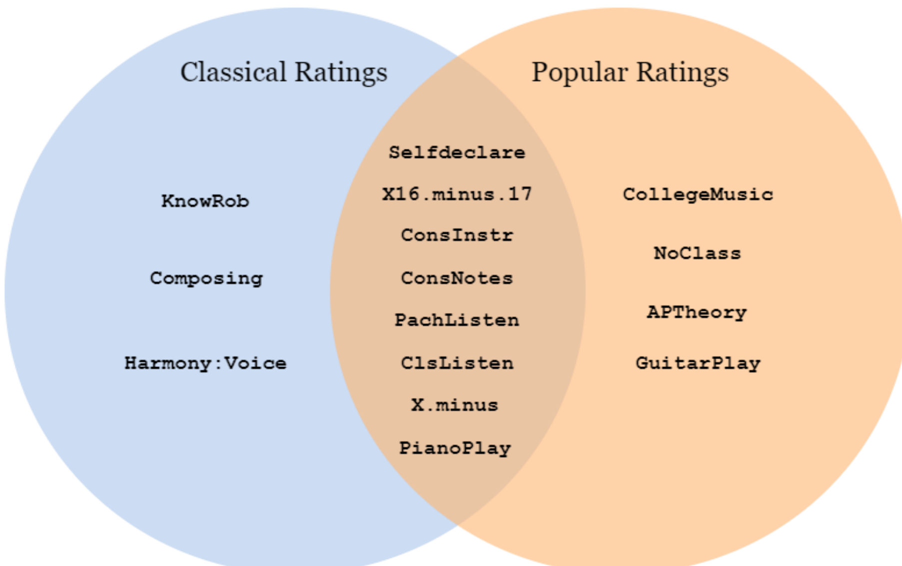

Figure 3: Venn Diagram of Model-Improving Fixed Effects

After back-fitting fixed effects and forward-fitting random effects, the automated method chooses quite a few fixed effects as predictors for both musical ratings. This may be due to the fact that imputation has lowered the variability within the predictors, making it easier for variables to become significant predictors. We should keep this issue in mind while interpreting the fixed effects.

While selecting random effects, the automated methods coincidentally chose the same ones for both models: the random intercept and the random slopes for Instrument and Harmony, all varying by subject, are included. In addition, for both models, we assume correlation between the random effects of Instrument and of Harmony. After obtaining the models, the residuals are again checked to examine model appropriateness. We looked at the binned residual pots, the plot of marginal residuals vs. fitted marginal values, and the normal QQ plots for both conditional and random effect residuals. In general, the residuals all display a reasonable pattern suggesting satisfaction of underlying model assumptions. Thus, our models are adequate in predicting the musical ratings.

[6]Please refer to Appendix D for the codes for developing the full model and the summary of the full model.

We first note that not all predictors in the Figure 3 is significant, although the automated method selected them to be included in the final model. Referring back to Figure 3, the following variables affect both classical and popular ratings:

- `Selfdeclare`: almost all levels in this variable are significant in predicting both classical and popular ratings. Interestingly, we found that self-identification as musician is actually negatively associated with classical ratings, and positively associated with popular ratings. In particular, level 5 in `Selfdeclare` has the strongest associated with music ratings.
- `X16.minus.17`: this variable is both positively associated with higher classical ratings (not significant) and higher popular ratings (significant).
- `ConsInstr`: for this variable, all levels except 5 are significant. For lower levels 1 and 2 are negatively associated with classical ratings and popular ratings, while higher levels 3 and 4 are positively associated with both higher classical ratings and lower popular ratings. In other words, the more the subject concentrate on the Instrument, the more they are likely to rate the music as classical but not popular, and vice versa.
- `ConsNotes`, `PachListen`, `ClsListen`: The behavior of these three predictors are highly consistent. All levels in these three predictors are significantly positively associated with higher classical ratings, and significantly negatively associated with lower popular ratings.
- `X.minus`: about half of the levels in this predictor are significant. In general, X.minus is negatively associated with higher classical ratings, and significantly positively associated with lower popular ratings.
- `PianoPlay`: while the first level of this predictor is positively associated with both classical and popular ratings, the second level and the fourth level has the opposite effect: negatively associated with classical ratings while positively associated with popular ratings.

In addition, there are indeed differences in the factors that drive popular and classical ratings. For the best classical model including the additional covariates, we found that level 1 of `KnowRob` is positively associated with higher classical ratings, while level 5 of `KnowRob` is negatively associated with higher classical ratings. We note that neither of the results are significant. In addition, we found that level 1 and 4 in `Composing` is significantly associated with higher classical ratings, level 2 is significantly associated with lower classical ratings. Lastly, the interaction between harmony and voice affects only classical ratings as well. The levels `HarmonyI-V-VI:Voicepar3rd` and `HarmonyIV-I-V:Voicepar3rd` are negatively and positively associated with classical ratings respectively, and both are significant. All the above interpretations are made by comparing to the baseline level of each predictor.

On the other hand, a completely different set of variables only impact popular ratings. We found that `CollegeMusic` and all levels in `NoClass` are negatively associated with popular ratings (about half of the levels are significant), while `APTheory` and all levels in `GuitarPlay` are positively associated with popular ratings (most levels are significant). For the first two predictors, we might suspect that both classes focus on classical music but not popular music, thus taking music classes in college might lead the subjects to be biased more to classical music. On other hand, since guitar is an essential component in popular music, it is reasonable that playing more guitar is positively associated with higher popular ratings.

3.3 The Influence of Being a Musician vs. a Non-musician for Classical Ratings

Finally, in this section, we address the hypothesis that self-identification as musician may have an influence on classical ratings. Based on the classical model developed in the previous section, we consider the interactions between the two musician variables and all other fixed effects in the model, and selected variables using stepwise methods and BIC values[7]. Spoiler alert: we found that while some interactions are not sensitive to where we dichotomize, some other interactions are only significant when we dichotomize in between specific levels. Next, let's denote the model with dichotomy between level 2 and level 3 as `mod_d23`, and the model with dichotomy between level 3 and level 4 as `mod_d34` for summarizing the results.

Below are part of the summary for `mod_d23`, containing all interactions selected by the automated method:

```
HarmonyI-V-IV:musician231  0.04781    0.25033   0.191 0.848541
HarmonyI-V-VI:musician231  1.20197    0.25033   4.802 1.67e-06 ***
HarmonyIV-I-V:musician231  0.06747    0.25011   0.270 0.787367
musician231:X16.minus.17  -0.32920    0.03432  -9.592  < 2e-16 ***
musician231:ClsListen      0.33697    0.06499   5.185 2.34e-07 ***
musician231:Composing     -0.57720    0.08541  -6.758 1.74e-11 ***
musician231:PianoPlay     -0.23074    0.05749  -4.014 6.16e-05 ***
```

Below are part of the summary for `mod_d34`, containing all interactions selected by the automated method:

```
HarmonyI-V-IV:musician341  0.10330    0.30279   0.341 0.733017
HarmonyI-V-VI:musician341  1.46471    0.30279   4.837 1.40e-06 ***
HarmonyIV-I-V:musician341  0.34448    0.30231   1.140 0.254602
musician341:X16.minus.17  -0.60777    0.06695  -9.078  < 2e-16 ***
musician341:KnowRob       -0.24250    0.06403  -3.787 0.000156 ***
musician341:X1990s2000s    0.34979    0.09774   3.579 0.000352 ***
musician341:X.minus       -0.35005    0.08583  -4.078 4.68e-05 ***
```

Selected by only `mod_d23`, the following variables have significant interactions with the musician variable. The interactions between the musician variable and the variables are sensitive in that these interactions are only statistically significant when we dichotomize `Selfdeclare` between level 2 and 3, indicating that the discrepancy in the extent to which people identify themselves as musicians is statistically significant between relatively lower levels.

- `ClsListen`: among the subjects who listen more to classical music, musicians (Selfdeclare $\geq$ 2) are significantly more likely to give higher classical ratings than the non-musicians (coef=0.33697).
- `Composing`: among the subjects who engage in more composing behavior, musicians (Selfdeclare $\geq$ 2) are significantly more likely to give lower classical ratings than the non-musicians (coef=-0.57720).

[7]Please refer to Appendix E for codes used to develop the model

- `PianoPlay`: among the subjects who play more piano, musicians (Selfdeclare ≥ 2) are significantly more likely to give lower classical ratings than the non-musicians (coef=-0.23074).

Selected only by `mod_d34`, the following variables have significant interactions with the musician variable. These interaction between the musician variable and the variables are sensitive in that these interactions are only statistically significant when we dichotomize `Selfdeclare` between level 3 and 4, indicating that the discrepancy in the extent to which people identify themselves as musicians is statistically significant between relatively higher levels.

- `KnowRob`: among the subjects who know Rob Paravonian's Pachelbel Rant well, musicians (Selfdeclare ≥ 3) are significantly more likely to give lower classical ratings than the non-musicians (coef=-0.24).

- `X1990s2000s`: among the subjects who listen more to pop and rock from the 90's and 2000's, musicians (Selfdeclare ≥ 3) are significantly more likely to give higher classical ratings than the non-musicians (coef=0.35).

- `X.minus`: among the subjects who have a larger preference over pop and rock from either 70's or 90's and 2000's, musicians (Selfdeclare ≥ 3) are significantly more likely to give lower classical ratings than the non-musicians (coef=-0.35).

Selected by both `mod_d23` and `mod_d34`, the following variables have significant interactions with the musician variable. The variables `Harmony` and `X16.minus.17` are not sensitive in where we dichotomize, since the interaction between the musician variable and them are significant for both dichotomy methods, indicating that discrepancies between level 2 and 3 and between level 3 and 4 in the extent to which people identify themselves as musicians would both cause significant differences in their ratings.

- `Harmony`: If the musical piece followed the harmonic motion I-V-VI, the musicians are significantly more likely to give higher Classical ratings than non-musicians (coef=1.20 for `mod_d23`, coef=1.46 for `mod_d34`). The interactions between the musician variable and other levels of Harmony are not significant.

- `X16.minus.17`: among the subjects with high score on the measure of a listener's ability to distinguish classical and popular music, the musicians are significantly more likely to give lower classical ratings than the non-musicians (coef=-0.33 for `mod_d23`, coef=-0.61 for `mod_d34`).

In conclusion, we confirmed the researchers' secondary hypothesis such that people who self-identify as musicians are indeed influenced by things that do not influence non-musicians. In particular, we found that the interactions with `ClsListen`, `Composing` and `PianoPlay` are significant and thus sensitive to dichotomy of Selfdeclare between level 2 and 3, whereas the interactions with `KnowRob`, `X1990s2000s` and `X.minus` are significant and thus sensitive to dichotomy of Selfdeclare between level 3 and 4. In addition, we also found that the interactions between `Harmony` and `X16.minus.17` are not sensitive to where we dichotomize: the interactions are significant in both kinds of dichotomy.

4 Discussion

The identification of music as classical or popular by subjects with varying musical expertise are found to be impacted by a variety of factors in a variety of ways. Out of the 23 variables from the dataset collected by Jimenez & Rossi (2012), we first investigated and compared the influence of Instrument, Harmony and Voice on musical ratings, then identified additional covariates that are useful for predicting classical and popular ratings respectively, and finally examined whether self-identification as musician may have an influence on classical ratings.

For the three main experimental factors, the mixed effects model shows that the statistical effect of the instrument used in the music on classical/popular ratings is the highest. For instruments, the classical ratings are highest when the instrument used in the musical stimuli is a string quartet, followed by piano, and lastly guitar. On the contrary, the popular ratings are highest when the instrument used in the musical stimuli is a guitar, followed by piano, and lastly string quartet - the complete opposite pattern when compared to the results of classical ratings. Intuitively, the result aligns with our common sense: whereas piano and string quartets are typical components in classical music, guitar has been widely used in popular music such as pop and rock. It is also worth noting that, instrument is found to be the only significant predictor of popular ratings. This result confirms one of the researchers' main hypotheses that instrument has the strongest influence on musical ratings.

In addition, the harmonic progression I-V-VI in the music is found to be strongly associated with higher classical ratings when compared to other harmonic progressions, confirming another researchers' hypothesis. Furthermore, the interaction between Harmony I-V-VI and parallel 3rd in Voice Leading is also found to be significant in predicting higher classical ratings. The researchers with both statistical knowledge and musical expertise may want to probe into this interesting relationship between harmony and voice leading and investigate what causes such a specific interaction to be significant in predicting higher classical ratings.

For the last experimental factor, voice leading, our model shows that none of its levels has significantly different impact on classical ratings. Although Voice is included in the final mixed effects model for classical ratings, we need to keep in mind that the neither of the automated methods for linear models or for mixed effects model included Voice as a predictor - this predictor is manually added back into the model simply because it is one of the three main design factors. Therefore, even though previous researches have shown that contrary motion in voice leading is frequently rated as classical, such pattern is not found in our model.

Apart from the three main experimental factors, some other covariates from the dataset are also found to be significant predictors of both music ratings. In particular, we found that the many levels in the variables `Selfdeclare`, `X16.minus.17`, `ConsInstr:`, `ConsNotes,`, `PachListen,`, `ClsListen:`, `X.minus` and `PianoPlay` significantly impact both musical ratings[8]. One interesting general pattern we noticed is that, if one level of a particular variable is found to be associated with higher classical ratings, it is highly likely that this level is associated with lower popular ratings at the same time. This pattern of having opposite signs suggests that, to some extent, the subjects implicitly assume that popular and classical musics are on two ends of the "musical spectrum". Although technically the subjects are allowed to rate a musical stimuli as high or low in both classical ratings and popular ratings, the results imply that musical stimuli rated high in classical

[8]Please refer back to Section 3.2 for details in how these variables affect the ratings (e.g. direction, significance, etc.)

ratings tend to have low popular ratings, and vice versa.

The mixed effects models also present a few variables that are significant in predicting only one kind of ratings. While the variables `KnowRob`, `Composing` and `Harmony:Voice` are statistically significant in predicting classical ratings only, the variables `CollegeMusic`, `NoClass`, `APTheory` and `GuitarPlay` are statistically significant in predicting popular ratings only. It is fascinating that the classical model chose `Composing` while the popular model chose `GuitarPlay`, because these two variables are highly correlated with a 0.68 correlation coefficient. Overall, this result confirms the speculation that there are indeed different factors that drive the classical and popular ratings respectively.

For all models used above, we included random effects that vary by subjects, which make our mixed effects models hierarchical. Up to this point, there have been four different models: predicting classical/popular ratings with three main experimental factors and predicting classical/popular with main factors and additional covariates. Using the automated method of back-fitting fixed effects and forward-fitting random effects, the chosen random effects turn out to be the same ones for all four models: a random intercept and two random slopes of Instrument and Harmony are included. All random effects vary by subject, and we assume that there is some correlation between the random effects of Instrument and the random effects of Harmony. In other words, the automated method consistently suggests that there are a considerable amount of variation in subjects' judgment overall, and the influence of instrument and the influence of harmony on musical ratings also substantially vary by subject. This outcome is quite reasonable, as the subjects in the experiment have varying levels of musical expertise.

Moving onto our last finding, we discovered that musicians and non-musicians are different in their identification of classical music. Recall that we dichotomized the `Selfdeclare` variable twice between different levels. By examining the selected interactions of the musician variable and other predictors, we found that the interactions with `Harmony` and `X16.minus.17` are not sensitive to where we dichotomize: these two interactions are significant in both kinds of dichotomy. On the other hand, some variables are sensitive in where we dichotomize: interactions of the musician variable with `ClsListen`, `Composing` and `PianoPlay` are only significant when we dichotomize `Selfdeclare` between level 2 and 3, while the interactions of the musician variable with `KnowRob:`, `X1990s2000s:` and `X.minus` are significant when we dichotomize between level 3 and 4. It is implied by the results that this discrepancy in identification of classical music between musicians and non-musicians sometimes depend on the variable and can only be significant when we dichotomize between specific levels.

We take note of a few caveats of the analysis and give suggestions on improving the current model and some directions for future research. Firstly, many variables given in the dataset measures similar information about the musical expertise of the subjects, thus leading to the potential issue of collinearity. For example, variables `CollegeMusic`, `NoClass` and `APTheory` all contains information on whether the subject has taken a college-level music class. Thus during the model building process, we tried to avoid collinearity in the predictors as much as possible by checking the variance inflation factors and the correlations between variables when applicable. Future researches can define a few general aspects of musical expertise and for each aspect use several related variables as a combined effect for prediction. This will not only shrink the number of the covariates, but also make the model interpretation more precise.

The second caveat in our analysis is that single median imputation is used to fill in many missing values in the dataset. As a result, some previously not significant variables might become significant in our model due to reduced variability, and this may be the reason why the model in Section 3.2 is quite large. It is recommended for future researchers to perform multiple imputation using regression estimation so as to improve the model.

Lastly, recall that all subjects in this experiment are undergraduates from Pittsburgh University. Since this group of subjects is not representative of the general public and voluntary response bias may exist, a more comprehensive analysis on how musical identification is affected can be done with a more representative sample of subjects, so that the prediction model is more conclusive and applicable to the general public.

With the admitted drawbacks in the models, evidence from various types of residual plots has suggested that our models are reasonably adequate in predicting classical and popular ratings. In summary, we found that while guitar is associated with higher popular ratings, string quartets and Harmony I-V-VI are associated with higher classical ratings. In addition, various aspects in musical knowledge (i.e. additional covariates in the dataset) and self-identification as a musician also have significant impact on musical ratings.

References

Jimenez I., Rossi V. (2012). The Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music. Pittsburhg, PA.

Prieto-Rodríguez, J., Fernández-Blanco, V. (2000). Are popular and classical music listeners the same people?. *Journal of Cultural Economics, 24*(2), 147-164.

Halpern, A. R., Bartlett, J. C. (2010). Memory for melodies. In *Music perception* (pp. 233-258). Springer, New York, NY.

R Core Team (2017), *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Appendix

A. CRAN Libraries Used for Analysis

The following libraries from the CRAN repository is used for performing all data analysis presented in the paper.

```
library(skimr)
library(ggcorrplot)
library(tidyverse)
library(ggplot2)
library(arm)
library(car)
library(LMERConvenienceFunctions)
library(RLRsim)
library(randomForest)
```

B. Section 3: Codes for Data Cleaning and EDA

Skimming data, deleting columns and renaming column names:

```
ratings = read.csv("ratings.csv", header = TRUE)
summary(ratings)
skim(ratings)
colnames(ratings)[colnames(ratings)=="X1990s2000s.minus.1960s1970s"] <- "X.minus"
ratings = ratings[, -26][, -25][, -24][, -11][, -1]
ind_cla = which(is.na(ratings$Classical)==TRUE)
ind_pop = which(is.na(ratings$Popular)==TRUE)
# they are the same rows - so we just delete once
ratings = ratings[-c(ind_cla), ]
```

Median imputations for missing values:

```
ratings$ConsNotes[which(is.na(ratings$ConsNotes))] <-
        median(ratings$ConsNotes, na.rm = TRUE)
ratings$PachListen[which(is.na(ratings$PachListen))] <-
        median(ratings$PachListen, na.rm = TRUE)
ratings$ClsListen[which(is.na(ratings$ClsListen))] <-
        median(ratings$ClsListen, na.rm = TRUE)
ratings$KnowRob[which(is.na(ratings$KnowRob))] <-
        median(ratings$KnowRob, na.rm = TRUE)
ratings$KnowAxis[which(is.na(ratings$KnowAxis))] <-
        median(ratings$KnowAxis, na.rm = TRUE)
ratings$X1990s2000s[which(is.na(ratings$X1990s2000s))] <-
        median(ratings$X1990s2000s, na.rm = TRUE)
ratings$X.minus[which(is.na(ratings$X.minus))] <-
        median(ratings$X.minus, na.rm = TRUE)
```

```
ratings$CollegeMusic[which(is.na(ratings$CollegeMusic))] <-
        median(ratings$CollegeMusic, na.rm = TRUE)
ratings$NoClass[which(is.na(ratings$NoClass))] <-
        median(ratings$NoClass, na.rm = TRUE)
ratings$APTheory[which(is.na(ratings$APTheory))] <-
        median(ratings$APTheory, na.rm = TRUE)
ratings$Composing[which(is.na(ratings$Composing))] <-
        median(ratings$Composing, na.rm = TRUE)
```

Histograms, correlation plots for continuous variables and categorical variables with more than 2 numeric levels:

```
ratings_nums = ratings[, c("APTheory", "ClsListen", "CollegeMusic", "Composing",
"ConsNotes", "GuitarPlay", "KnowAxis", "KnowRob", "NoClass", "OMSI", "PachListen",
"PianoPlay", "Selfdeclare", "X.minus", "X1990s2000s", "Classical", "ConsInstr",
"Popular", "X16.minus.17")]

par(mfrow=c(3,3))
for (i in names(ratings_nums)[-1]) {
  hist(ratings_nums[,i], main=i, xlab = i)
}

par(mfrow=c(1,1))
corr = round(cor(ratings_nums),2)
ggcorrplot(corr, type="lower", title="Variable Correlation",
           colors = c("#6D9EC1","white","#E46726"), show.diag = TRUE,
           lab=TRUE, lab_size = 3, insig = c("pch","blank"))
```

Based on the results from the codes above, we perform a logarithm transformation on the skewed variable `OMSI`:

```
ratings$OMSI_log = log(ratings$OMSI)
```

Box plots for `Classical` and `Popular` ratings:

```
ratings_cat = ratings[,-24][, -23][,-22][,-17][,-6]
# Classical
par(mfrow = c(3, 4))
for (i in 1:ncol(ratings_cat)) {
  plot(x = as.factor(ratings_cat[, i]), y = ratings$Classical,
       xlab = colnames(ratings_cat)[i], ylab = "Classical")
}
# Popular
par(mfrow = c(3, 4))
for (i in 1:ncol(ratings_cat)) {
  plot(x = as.factor(ratings_cat[, i]), y = ratings$Popular,
       xlab = colnames(ratings_cat)[i], ylab = "Popular")
}
```

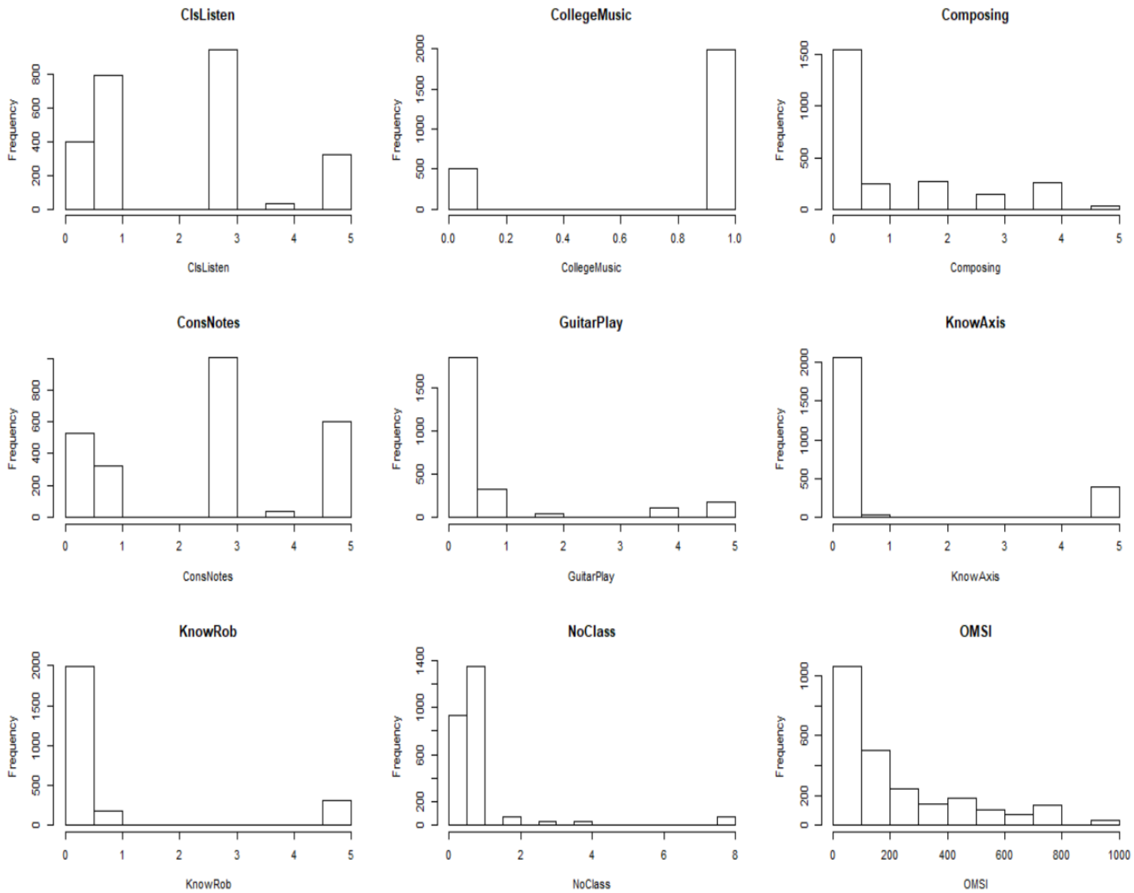

Figure 4: Sample Histograms of Variable Distribution. Jimenez & Rossi (2012).

C. Section 3.1: Codes for Model Selection

Sample model selection process for classical ratings are presented below. The process for selecting a model for popular ratings is identical, except the response variable is changed to be `Popular`.

```
# FIXED EFFECTS
lm1 = lm(Classical ~ (Instrument + Harmony + Voice)^3, data = ratings)
stepAIC(lm1, direction="both", k=2)
stepAIC(lm1, direction="both", k=log(nrow(ratings)))
lmAIC = lm(Classical ~ Instrument + Harmony + Voice + Harmony:Voice, data=ratings)
lmBIC = lm(Classical ~ Instrument + Harmony + Voice, data=ratings)
anova(lmAIC, lmBIC)
# selected predictors: ‘Instrument‘, ‘Harmony‘, ‘Voice‘, and ‘Harmony:Voice‘

# RANDOM EFFECTS
# best subsets
lmer_i = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
```

```
        (1|Subject) + (Instrument|Subject), data=ratings, REML=FALSE,
    control=lmerControl(optimizer="bobyqa"))
lmer_h = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Harmony|Subject), data=ratings, REML=FALSE,
    control=lmerControl(optimizer="bobyqa"))
lmer_v = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) +(Voice|Subject), data=ratings, REML=FALSE,
    control=lmerControl(optimizer="bobyqa"))

# two random effects, assume no correlation
lmer_ih = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument|Subject)+(Harmony|Subject), data=ratings,
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))
lmer_iv = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument|Subject)+(Voice|Subject), data=ratings,
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))
lmer_hv = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Harmony|Subject) + (Voice|Subject), data=ratings,
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))

# two random effects, assume correlation
lmer_ih_c = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument+Harmony|Subject), data=ratings, REML=FALSE,
        control=lmerControl(optimizer="bobyqa"))
lmer_iv_c = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument+Voice|Subject), data=ratings, REML=FALSE,
        control=lmerControl(optimizer="bobyqa"))
lmer_hv_c = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Harmony+Voice|Subject), data=ratings, REML=FALSE,
        control=lmerControl(optimizer="bobyqa"))

# three random effects
lmer_ihv = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument|Subject)+(Harmony|Subject)+(Voice|Subject),
        data=ratings, REML=FALSE, control=lmerControl(optimizer="bobyqa"))
lmer_ihv_c = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument+Harmony+Voice|Subject),
        data=ratings, REML=FALSE, control=lmerControl(optimizer="bobyqa"))

# the last few combinations
lmer_ih_c_v = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument+Harmony|Subject)+(Voice|Subject), data=ratings,
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))
lmer_iv_c_h = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Instrument+Voice|Subject)+(Harmony|Subject), data=ratings,
```

```
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))
lmer_vh_c_i = lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony*Voice +
        (1|Subject) + (Harmony+Voice|Subject)+(Instrument|Subject), data=ratings,
        REML=FALSE, control=lmerControl(optimizer="bobyqa"))

# compare AIC/BIC values for all subsets
anova(lmer_h, lmer_v, lmer_hv, lmer_iv, lmer_ih, lmer_iv_c, lmer_ih_c, lmer_hv_c,
      lmer_ihv, lmer_ihv_c, lmer_ih_c_v, lmer_iv_c_h,lmer_vh_c_i)

# automated
summary(fitLMER.fnc(lmer1, ran.effects=c("(Harmony|Subject)",
        "(Instrument|Subject)", "(Voice|Subject)"), method="BIC"))
```

Checking residuals:

```
resid.marg_2 <- r.marg(lmer_ih_c)
resid.cond_2 <- r.cond(lmer_ih_c)
resid.reff_2 <- r.reff(lmer_ih_c)
fit.marg_2 <- yhat.marg(lmer_ih_c)
par(mfrow=c(2,2))
binnedplot(predict(lmer_ih_c), resid(lmer_ih_c))
plot(fit.marg_2, resid.marg_2, main = "Marginal Residuals")
abline(h=0)
lines(loess.smooth(fit.marg_2, resid.marg_2), col="red")
qqnorm(resid.cond_2, main="Conditional Residuals")
qqline(resid.cond_2)
qqnorm(resid.reff_2, main="Random Effect Residuals")
qqline(resid.reff_2)
```

D. Section 3.2: Codes for Adding Other Predictors

Sample model selection process for classical ratings are presented below. The process for selecting a model for popular ratings is identical, except the response variable is changed to be `Popular`.

```
# FIXED EFFECTS
lm_q3_fixed = lm(Classical ~ Harmony + Instrument + Voice + Harmony:Voice +
        Selfdeclare + X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
        KnowRob + KnowAxis + X1990s2000s + X.minus + CollegeMusic + NoClass +
        APTheory + Composing + PianoPlay + GuitarPlay + OMSI_log, data=ratings_fac)

# AIC: dropped CollegeMusic, NoClass, APTheory
stepAIC(lm_q3_fixed, direction="both", k=2)

# BIC: dropped Harmony:Voice, KnowAxis, CollegeMusic, NoClass, APTheory, GuitarPlay
stepAIC(lm_q3_fixed, direction="both", k=log(nrow(ratings_fac)))
```

We drop CollegeMusic, NoClass, and APTheory, KnowAxis, GuitarPlay. We keep Harmony:Voice. We also drop the OMSI_log column, since Selfdeclare is highly correlated with OMSI_log.

```
# RANDOM EFFECTS
lmer_q3_full = lmer(Classical ~ Harmony + Instrument + Voice + Harmony:Voice +
        Selfdeclare + X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
        KnowRob + X1990s2000s + X.minus + Composing + PianoPlay + (1|Subject),
        REML=FALSE, data=ratings_fac, control=lmerControl(optimizer="bobyqa"))

fitLMER.fnc(lmer_q3_full, ran.effects=c("(Instrument|Subject)", "(Harmony|Subject)",
        "(Voice|Subject)"), method="AIC", keep.single.factors = TRUE)
```

Summary of final classical model:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ Harmony + Instrument + Voice + Selfdeclare + X16.minus.17 +
##     ConsInstr + ConsNotes + PachListen + ClsListen + KnowRob +
##     X1990s2000s + X.minus + Composing + PianoPlay + (1 | Subject) +
##     (Instrument + Harmony | Subject) + Harmony:Voice
##    Data: ratings_fac
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   9907.2  10425.3  -4864.6   9729.2     2404
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7281 -0.6019  0.0146  0.5758  6.1945
##
## Random effects:
##  Groups    Name             Variance  Std.Dev.  Corr
##  Subject   (Intercept)      2.549e-14 1.596e-07
##  Subject.1 (Intercept)      2.158e+00 1.469e+00
##            Instrumentpiano  1.644e+00 1.282e+00 -0.34
##            Instrumentstring 3.519e+00 1.876e+00 -0.73  0.65
##            HarmonyI-V-IV    3.927e-02 1.982e-01  0.36 -0.63 -0.49
##            HarmonyI-V-VI    1.582e+00 1.258e+00 -0.29 -0.27 -0.42  0.12
##            HarmonyIV-I-V    2.049e-03 4.527e-02 -0.15 -0.57  0.14 -0.02
##  Residual                   2.402e+00 1.550e+00
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)                0.92589    0.99755   0.928
## HarmonyI-V-IV              0.15537    0.15409   1.008
```

```
## HarmonyI-V-VI              1.13615    0.21385   5.313
## HarmonyIV-I-V             -0.13384    0.15171  -0.882
## Instrumentpiano            1.37177    0.17140   8.003
## Instrumentstring           3.12799    0.23671  13.214
## Voicepar3rd               -0.27193    0.15167  -1.793
## Voicepar5th               -0.23555    0.15203  -1.549
## Selfdeclare2              -1.14599    0.36765  -3.117
## Selfdeclare3              -0.82445    0.38125  -2.163
## Selfdeclare4              -1.44502    0.41908  -3.448
## Selfdeclare5              -4.56595    0.99972  -4.567
## Selfdeclare6              -1.63245    1.12934  -1.445
## X16.minus.17               0.05184    0.04320   1.200
## ConsInstr1                -0.79416    0.41740   5.373
## ConsInstr2                -5.81214    1.26975  -7.577
## ConsInstr3                 1.07484    0.37375   5.876
## ConsInstr4                 4.84272    0.88045   6.500
## ConsInstr5                 0.08397    0.40887   0.205
## ConsNotes1                 1.12843    0.45427   2.484
## ConsNotes3                 1.93502    0.37522   5.157
## ConsNotes4                 1.66752    0.77536   2.151
## ConsNotes5                 1.71399    0.48289   3.549
## PachListen1                5.91330    1.14094   5.183
## PachListen2                1.91374    0.75610   2.531
## PachListen3                2.22889    0.84691   2.632
## PachListen5                3.64178    0.80035   4.550
## ClsListen1                 1.48994    0.41148   3.621
## ClsListen3                 2.51728    0.46111   5.459
## ClsListen4                 3.73186    1.14995   3.245
## ClsListen5                 4.37006    0.68431   6.386
## KnowRob1                   0.96291    0.55209   1.744
## KnowRob5                  -0.09300    0.34138  -0.272
## X1990s2000s2              -4.29975    0.79719  -5.394
## X1990s2000s3              -3.42357    0.70009  -4.890
## X1990s2000s4              -2.14307    0.65479  -3.273
## X1990s2000s5              -2.32690    0.52706  -4.415
## X.minus-3                 -5.61494    1.22675  -4.577
## X.minus-2                 -5.47472    1.02300  -5.352
## X.minus0                  -0.38205    0.37177  -1.028
## X.minus1                  -1.46750    0.83594  -1.756
## X.minus2                  -1.04317    0.38504  -2.709
## X.minus3                  -0.99960    0.36461  -2.742
## X.minus4                  -0.40612    0.64255  -0.632
## Composing1                 1.39266    0.35076   3.970
## Composing2                -1.14847    0.43833  -2.620
## Composing3                -0.77221    0.54067  -1.428
```

```
## Composing4                1.22328    0.44576   2.744
## PianoPlay1                0.78567    0.30713   2.558
## PianoPlay2               -2.50891    1.08154  -2.320
## PianoPlay4               -1.19070    0.43084  -2.764
## PianoPlay5               -0.01298    0.50759  -0.026
## HarmonyI-V-IV:Voicepar3rd -0.36433    0.21496  -1.695
## HarmonyI-V-VI:Voicepar3rd -0.67997    0.21519  -3.160
## HarmonyIV-I-V:Voicepar3rd  0.48754    0.21486   2.269
## HarmonyI-V-IV:Voicepar5th -0.19046    0.21538  -0.884
## HarmonyI-V-VI:Voicepar5th -0.42536    0.21513  -1.977
## HarmonyIV-I-V:Voicepar5th  0.07349    0.21474   0.342
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 3 columns / coefficients
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Checking residuals:

```
resid.marg_3 <- r.marg(lmer_q3_final)
resid.cond_3 <- r.cond(lmer_q3_final)
resid.reff_3 <- r.reff(lmer_q3_final)
fit.marg_3 <- yhat.marg(lmer_q3_final)
par(mfrow=c(2,2))
binnedplot(predict(lmer_q3_final), resid(lmer_q3_final))
plot(fit.marg_3, resid.marg_3, main = "Marginal Residuals")
abline(h=0)
lines(loess.smooth(fit.marg_3, resid.marg_3), col="red")
qqnorm(resid.cond_3, main="Marginal Residuals")
qqline(resid.cond_3)
qqnorm(resid.reff_3, main="Random Effect Residuals")
qqline(resid.reff_3)
```

E. Section 3.3: Codes for Creating Musician Variable

Two dichotomies

```
ratings_mus = ratings

# between 2 and 3
ratings_mus <- ratings_mus %>%
        mutate(musician23 = ifelse(Selfdeclare %in% c(1,2), 0, 1))
ratings_mus$musician23 = as.factor(ratings_mus$musician23)

# between 3 and 4
ratings_mus <- ratings_mus %>%
        mutate(musician34 = ifelse(Selfdeclare %in% c(1,2,3), 0, 1))
ratings_mus$musician34 = as.factor(ratings_mus$musician34)
```

Model selection procedure for two models dichotomized at different levels

```
# two full models
lm_q4_d23 = lm(Classical ~ Harmony + Instrument + Voice + Harmony:Voice + musician23 +
        X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen + KnowRob +
        X1990s2000s + X.minus + Composing + PianoPlay + Harmony:musician23 +
        Instrument:musician23 + Voice:musician23 + Harmony:Voice +
        X16.minus.17:musician23 + ConsInstr:musician23 + ConsNotes:musician23 +
        PachListen:musician23 + ClsListen:musician23 + KnowRob:musician23 +
        X1990s2000s:musician23 + X.minus:musician23 + Composing:musician23 +
        PianoPlay:musician23, data=ratings_mus)

lm_q4_d34 = lm(Classical ~ Harmony + Instrument + Voice + Harmony:Voice + musician34 +
        X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen + KnowRob +
        X1990s2000s + X.minus + Composing + PianoPlay + Harmony:musician34 +
        Instrument:musician34 + Voice:musician34 + Harmony:Voice +
        X16.minus.17:musician34 + ConsInstr:musician34 + ConsNotes:musician34 +
        PachListen:musician34 + ClsListen:musician34 + KnowRob:musician34 +
        X1990s2000s:musician34 + X.minus:musician34 + Composing:musician34 +
        PianoPlay:musician34, data=ratings_mus)

# AIC selection
summary(stepAIC(lm_q4_d23, k = 2, direction = "both"))
# HarmonyI-V-IV:musician231     0.04560    0.24833   0.184  0.85432
# HarmonyI-V-VI:musician231     1.19914    0.24833   4.829 1.46e-06 ***
# HarmonyIV-I-V:musician231     0.06742    0.24811   0.272  0.78583
# Instrumentpiano:musician231  -0.61336    0.21581  -2.842  0.00452 **
# Instrumentstring:musician231 -0.70007    0.21413  -3.269  0.00109 **
# musician231:X16.minus.17     -0.32831    0.03497  -9.388  < 2e-16 ***
# musician231:ConsNotes         0.12068    0.05993   2.014  0.04414 *
# musician231:PachListen       -0.20847    0.11963  -1.743  0.08152 .
# musician231:ClsListen         0.38523    0.07692   5.008 5.88e-07 ***
# musician231:KnowRob          -0.13769    0.07147  -1.926  0.05416 .
# musician231:Composing        -0.57110    0.09063  -6.301 3.49e-10 ***
# musician231:PianoPlay        -0.25217    0.06157  -4.096 4.35e-05 ***

summary(stepAIC(lm_q4_d34, k = 2, direction = "both"))
# HarmonyI-V-IV:musician341     0.10552    0.29999   0.352 0.725055
# HarmonyI-V-VI:musician341     1.46701    0.29999   4.890 1.07e-06 ***
# HarmonyIV-I-V:musician341     0.34557    0.29950   1.154 0.248686
# Instrumentpiano:musician341  -0.46446    0.26075  -1.781 0.074994 .
# Instrumentstring:musician341 -0.94717    0.25852  -3.664 0.000254 ***
# musician341:X16.minus.17     -0.74387    0.07691  -9.672  < 2e-16 ***
# musician341:ConsInstr        -0.28862    0.09940  -2.904 0.003721 **
# musician341:PachListen       -0.91613    0.29898  -3.064 0.002206 **
# musician341:ClsListen         0.54876    0.13680   4.012 6.21e-05 ***
```

```
# musician341:KnowRob          -0.46909    0.08796  -5.333 1.05e-07 ***
# musician341:X1990s2000s       0.65456    0.12576   5.205 2.10e-07 ***
# musician341:X.minus          -0.62907    0.12435  -5.059 4.53e-07 ***

# BIC selection
summary(stepAIC(lm_q4_d23, k = log(nrow(ratings_mus)), direction = "both"))
# HarmonyI-V-IV:musician231  0.04781    0.25033   0.191 0.848541
# HarmonyI-V-VI:musician231  1.20197    0.25033   4.802 1.67e-06 ***
# HarmonyIV-I-V:musician231  0.06747    0.25011   0.270 0.787367
# musician231:X16.minus.17  -0.32920    0.03432  -9.592  < 2e-16 ***
# musician231:ClsListen      0.33697    0.06499   5.185 2.34e-07 ***
# musician231:Composing     -0.57720    0.08541  -6.758 1.74e-11 ***
# musician231:PianoPlay     -0.23074    0.05749  -4.014 6.16e-05 ***

summary(stepAIC(lm_q4_d34, k = log(nrow(ratings_mus)), direction = "both"))
# HarmonyI-V-IV:musician341  0.10330    0.30279   0.341 0.733017
# HarmonyI-V-VI:musician341  1.46471    0.30279   4.837 1.40e-06 ***
# HarmonyIV-I-V:musician341  0.34448    0.30231   1.140 0.254602
# musician341:X16.minus.17  -0.60777    0.06695  -9.078  < 2e-16 ***
# musician341:KnowRob       -0.24250    0.06403  -3.787 0.000156 ***
# musician341:X1990s2000s    0.34979    0.09774   3.579 0.000352 ***
# musician341:X.minus       -0.35005    0.08583  -4.078 4.68e-05 ***
```