Classcial or Popular? - An Investigation of Music Genre Recognition

Yichun Shi*
yichunsh@andrew.cmu.edu

December 7, 2019

Abstract

We have investigated the relationship between instrument, harmonic motion, and voice leading of a music piece with listener's recognition of the piece as Classical and Popular genre. Data cleaning was performed, and mixed models were constructed based on automatic variable selection methods. Various statistical techniques including AIC, BIC, and partial-F tests were applied to evaluate the models. The result from our models suggests that the instrument used, harmonic motion of the song, the interaction between harmonic motion and voice leading, and the listener's mastery of piano are significantly associated with the listener's recognition of music as Classical. For Popular genre, the instrument used, and the interaction between harmonic motion and whether the listener self-identifies as musician are important predictors of the genre rating given by the listener. For both genres, the instrument used has the strongest influence for people's genre recognition, with piano and string more associated with Classical and guitar more associated with Popular. Future researches should perform appropriate imputation to address the missing data problem of music ratings data set in order to derive model that avoids introducing bias while reflects a more complete picture of the original data set. Objective, quantitative variables such as the OMSI scores should be used to define musician versus non-musicians as opposed to using people's self-rating in order to decrease impact the personal bias on the predictor, and generate conclusions that gives more insight on how musicians use information in the music pieces differently to determine the music genre.

1 Introduction

Music genre is a category convention that allows peoples to identify and group music pieces that shares a same set of conventions (Wikipedia.org, 2019). However, such criteria to classify music into genres are often subjective and controversial, hence leading to many different frameworks to assess the genre of a music piece. Traditional approaches of music genre classification include analyzing social, organizational, and symbolic attributes of the music piece (Lena et al, 2008). Modern approaches of music genre recognition is heavily focused on classification modeling methods, such as analyzing audio spectrograms using Convolutional Neural Networks (Oramas et al, 2018).

*Department of Statistics & Data Science, Carnegie Mellon University

In this study, we will be focusing on how listeners process and use three types of information presented in a musical sample, namely the instrument used to perform the piece, harmonic motion of the piece, and voice leading, to recognize the music genre as either Classical or Popular. Specifically, we are looking to address the following research questions:

- Which of the three experimental factor, or combination of factors, has the strongest influence on the audience's recognition of music genre?
- Among the levels of harmonic motions, does I-V-Vi have the strongest association with Classical ratings? Does it seem to matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits?
- Among the levels of voice leading, does contrary motion have the strongest association with classical ratings?
- Are there differences in the way that musicians and non-musicians identify classical music?
- Besides the three main experiment factors, what are the other factors that are associated with audience's recognition of music genre? Do these factors differ for Popular and Classical genre?

The data was collected by in Jimenez and Rossi. (2013).

2 Methods

2.1 Data

We will be examining the influence of instrument, harmonic motion and voice leading on audience's identification of music as Classical or Popular genre. All data used in this study were collected by Ivan Jimenez and Vincent Rossi in 2012. 70 undergraduate students at University of Pittsburgh were recruited as test subjects for this study. In the experiment, the test subjects were presented with 36 musical stimuli, and then were asked to rate these pieces on two scales:

- How classical does the music sound (1 to 10, 1=not at all, 10=very classical sounding)
- How popular does the music sound (1 to 10, 1=not at all, 10=very popular sounding).

Listeners were told that a piece could be rated as both classical and popular, neither classical nor popular, or mostly classical and not popular (or vice versa), so that the scales should have functioned more or less independently.

The data set contains 26 variables, with 2520 observations in total. Table 1 displays the definition of each variable. The data set `ratings.csv` can be retrieved on canvas course website of 36617 Applied Linear Models at Carnegie Mellon University.

As there exist variables in the data set with missing data, we have used the skimr package in R to investigate the amount of missing entries. For those variables with large amount of missingness, we have discarded them completely. For those variables with only small amount of missingness, we have omitted the rows with missing entries so that there are still a reasonable amount of data observations to work with when fitting the regression models.

The data set is consisted of continuous variables, ordinal variables, and dummy variables. We have investigated on the variables via histograms to check the distribution and range of the values. For some variables that should only take integers as inputs yet have decimal points, we have rounded them to the closest integers. For all variables that are defined as ordinal or dummy, we have treated them as factors in our models. In this way we were able to capture the specific differences between relationship of each level of the variable with the outcome variable when performing regression. A detailed steps for the EDA, data cleaning and transformation can be found on page 1 to 7 of the Appendix section.

2.2 Identify Impact on Genre Recognition among Instrument, Harmonic Motion, and Voice Leading

In order to construct a well-specified model that examines the relationship between our three main experiment factors and listener's judgement on music genre as Classical or Popular, we have not only constructed model with the three factors, but also have performed a careful selection on the interaction terms on the three main experiment factors. To achieve this, we have first purposed two regression models (one with *Classical* as outcome variable, the other with *Popular* as outcome variable) using the lm() function in R with all possible degrees of interactions between the three main experiment factors, then applied the automatic variable selection based on AIC value via the stepAIC() function from library MASS in R. A partial F-test was performed using the purposed reduced model from stepAIC() and the full model to ensure that the reduced model is well-specified.

We are also interested in finding whether the contrary motion has the strongest impact to rating as classical among all voice leading types. To achieve this, we have specifically make the function in R language to estimate the coefficients for all levels of $Voice$ when structuring the model for $Classical$, rather than taking $Contrary$ as the reference category. We have compared the size of the coefficients for each level of $Voice$ to check which level has the strongest influence on the outcome variable.

Additionally, one of our research questions is to discover whether among all harmonic motion types if the I-V-VI progression has the strongest association with rating as classical, and whether such association also is related to whether the subject is is familiar with the Pachelbel rants or comedy bits. To address this, we have investigated the significance and coefficient value of the I-V-VI level in the model. We have also tested whether adding interaction term of $KnowRob$ and $Harmony$, $KnowAxis$ and $Harmony$ is necessary via partial F-tests.

As the experiment was designed to have repeated-measures on the test subjects, we have investigated whether random intercept and random slopes on the three main experiment factors were needed in our model. For testing the random intercept, we have first added the random intercept to both Popular and Classical models, then checked whether the AIC and BIC values were reduced by adding the intercept. For the random slopes, we have first added all possible random slopes to the two models, then applied the fitLMER.fnc() function from library LMERConvenienceFunctions with three back-fitting methods (BIC, t-test, log ratio test) to the full model. With the reduced models generated using these three approaches, we manually check if there exist any random slope terms that were agreed by at least two out of three approaches of fitLMER.fnc(). For those terms that were agreed, we added them back to the model, and check if both AIC and BIC values dropped, which indicates an improvement of the model quality. The detailed steps can be found on page 11 to 15 for Classical, and Finally, we have investigated the residual plots to see whether the model assumptions are met.

Variable Number	Variable Name	Description
1	Classical	How classical does the stimulus sound?
2	Popular	How popular does the stimulus sound?
3	Subject	Unique subject ID
4	Harmony	Harmonic Motion (4 levels)
5	Instrument	Instrument (3 levels)
6	Voice	Voice Leading (3 levels)
7	Selfdeclare	Are you a musician? (1-6, 1=not at all)
8	OMSI	Score on a test of musical knowledge
9	X16.minus.17	Auxiliary measure of listener's ability to distinguish classical vs popular music
10	ConsInstr	How much did you concentrate on the instrument while listening (0-5, 0=not at all)
11	ConsNotes	How much did you concentrate on the notes while listening? (0-5, 0=not at all)
12	Instr.minus.Notes	Difference between prev. two variables
13	PachListen	How familiar are you with Pachelbel's Canonin D (0-5, 0=not at all)
14	ClsListen	How much do you listen to classical music? (0-5, 0=not at all)
15	KnowRob	Have you heard Rob Paravonian's Pachelbel Rant(0-5, 0=not at all)
16	KnowAxis	Have you heard Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music?(0-5, 0=not at all)
17	X1990s2000s	How much do you listen to pop and rock from the 90's and 2000's? (0-5, 0=not at all)
18	X1990s2000s. minus.1960s1970s	Difference between previous variable and a similar variable referring to 60'sand 70's pop and rock.
19	CollegeMusic	Have you taken music classes in college(0=no, 1=yes)
20	NoClass s	How many music classes have you taken?
21	APTheory	Did you take AP Music Theory class in High School (0=no, 1=yes)
22	Composing	Have you done any music composing (0-5,0=not at all)
23	PianoPlay	Do you play piano (0-5, 0=not at all)
24	GuitarPlay	Do you play guitar (0-5, 0=not at all)
25	X1stInstr	How proficient are you at your first musical instrument (0-5, 0=not at all)
26	X2stInstr	How proficient are you at your second musical instrument (0-5, 0=not at all)

Table 1: Variables in the music data set

2.3 Impact of Self-identification as Musician on Genre Recognition

As we are interested in analyzing whether the subjects identify themselves as musician is an important predictor for their rating of music genre, we have first dichotomized the variable *Selfdeclare* with various cut-off points (at 2, 3, 4), and used histograms to investigate the cut-off that would yield a roughly 50-50 split of the data. Then, we have tested whether adding the dichotomized *Selfdeclare* variable into the best model from section 2.2 is preferred based on partial F-test result.

We are also interested in investigating whether such self-identification changes the relationship between the three main experiment factors (instrument, harmonic motion, and voice leading) and music genre rating. To achieve this, we have tested adding random effects regarding the dichotomized *Selfdeclare* variable to the best model from section 2.2. We have added the random intercept, then checked whether AIC and BIC have decreased from the model without random intercept. For the random slopes, we have first added all possible random slopes on the existing fixed effects in the best model from section 2.2, then applied the fitLMER.fnc() function with three back-fitting methods. Again we have checked if there are random effects where two out of the three approaches agree, and confirmed whether adding those terms would lead to smaller AIC and BIC. We have performed this same procedure with the $Selfdeclare$ variable dichotomized at different cut-offs to check whether the results change based on the cut-off used. Finally, we have investigated the residual plots to see whether the model assumptions are met.

2.4 Covariates for Genre Recognition

To check whether there are other important predictors for the rating of music genre as Classical and Popular, we have conducted further variable selection using automatic methods in R. Based on the best models for *Classical* and *Popular* found in section 2.2, we have first selected the fixed effects by creating a linear model using the lm() in R with the rest of the transformed variables from the data set (note that *X1stInstr* and *X1stInstr* were omitted in this process due to large amount of missing values) added, then used the stepAIC() function to perform step-wise model selection. With the suggested covariates from stepAIC(), we added them as fixed effects back to the best mixed model from Section 2.2, performed fitLMER.fnc() to check if there are fixed effects agreed by two out of three approaches. After adding those agreed fixed effect to the model, partial F-test was performed to check whether adding the new fixed effects were preferred as opposed to the original reduced model.

We are also interested in checking for the selected candidate covariates, if there should be a subject-based random effect added to those covariates in the model. To achieve this, we added random slopes to these fixed effects in our model, and performed fitLMER.fnc() to conduct random effect selection. We have picked those random slopes that were agreed by at least two approaches of fitLMER.fnc(), and confirmed whether the AIC and BIC values decreased with that model specification. Finally, we have investigated the residual plots to see whether the model assumptions are met.

3 Results

3.1 Data Cleaning and Transformation

We have performed data cleaning on the original data set to prepare for the modeling. As discussed in section 2.1, there exist some degrees of missingness for multiple variables in the data set. As the variable *X1stInstr* and *X1stInstr* have majority of the entries missing (around 1000 and 300 complete observations), we have decided not to use these two variables as candidates for covariate selection to avoid omitting too many incomplete observations in our regression model. The missgingness issue for variable *APTheory*, *ClsListen*,*CollegeMusic*, *Composing*,*ConsNotes*, *KnowAxis*, *KnowRob*, *NoClass*, *PachListen*, *X1990s2000s* were less concerning, as the amount of missing entries is relatively small (less than 300 entries out of 2520 observations). None of the three main experiment factors had missing data. Hence, we have omitted the observations with missing entries only when conducting covariate selection. A detailed steps for checking the missingness in our variable set, and omitting the missing entries can be found on page X of the Appendix section.

We have also applied transformations as below on the data set. The detailed steps can be found on page 8 to 9 in the Appendix section.

1. For variable *ConsInstr*, we have rounded the entries with decimal values to closest integer. By definition, this variable is ordinal and should only take values of 0 to 5. It is likely that there was mistakes made when entering those observations into the data set.

2. We have performed log transformation of variable *OMSI*. This is because the variable was highly right-skewed, and the range of value of the variable is quite large (0 to 2000) comparing to the rest of variables in our data set. Hence, we have performed log transformation to avoid getting misleading regression results due to the large range of this variable.

3. For variables $Selfdeclare$, $PachListen$, $ClsListen$, $ConsInstr$, $ConsNotes$, $X1990s2000s$, $CollegeMusic$, $KnowRob$, $KnowAxis$, $PianoPlay$, $GuitarPlay$, and $APTheory$, we have treated them as factors in our models via the as.factor() function in R language as by definition they are either ordinal or dummy variables.

3.2 Impact of Instrument, Harmonic Motion and Voice Leading on Music Genre Recognition

To construct a well-specified model that examines the impact of instrument, harmonic motion, and voice leading on audience's rating of music genre as Classical or Popular, we have first performed variable selection via stepAIC() function on the interaction terms of the three main effects. The result suggested that for the model of *Classical*, the interaction term *Harmony * Voice* should be included in the model, while for the model of *Popular* no interaction is needed. Detailed procedure in R language can be found on page 9 to 10 (for *Classical*) and page 27 to 29 (for *Popular*) of the Appendix section.

We have also constructed mixed models instead of simple linear regression as we found that for both the case of modeling *Classical* and *Popular*, adding random effects regarding subject would decrease the AIC and BIC values, suggesting improvements in model specification. Detailed steps of using the fitLMER.fnc() function to conduct random effect selection can be found on page 10 to 15 (for *Classical*) and page 28 to 29 (for *Popular*) of the Appendix section.

The final mixed models for *Classical* and *Popular* are presented as follows:

$$
\begin{aligned}
Classical_i &= \alpha_{0j[i]} + \alpha_{1j[i]}Instrument + \alpha_{2j[i]}Harmony + \beta_3 Voice + \beta_4 Harmony * Voice + \epsilon_i, \epsilon_i \sim_{iid} N(0, \sigma^2) \\
\alpha_{0j} &= \beta_0 + \eta_{0j}, \eta_{0j} \sim_{iid} N(0, \tau_0^2) \\
\alpha_{1j} &= \beta_1 + \eta_{1j}, \eta_{1j} \sim_{iid} N(0, \tau_1^2) \\
\alpha_{2j} &= \beta_2 + \eta_{2j}, \eta_{2j} \sim_{iid} N(0, \tau_2^2)
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
Popular_i &= \alpha_{0j[i]} + \alpha_{1j[i]}Instrument + \beta_2 Harmony + \beta_3 Voice + \epsilon_i, \epsilon_i \sim_{iid} N(0, \sigma^2) \\
\alpha_{0j} &= \beta_0 + \eta_{0j}, \eta_{0j} \sim_{iid} N(0, \tau_0^2) \\
\alpha_{1j} &= \beta_1 + \eta_{1j}, \eta_{1j} \sim_{iid} N(0, \tau_1^2)
\end{aligned} \tag{2}
$$

The residual analysis for Model 1 (see page 18 to 27 in Appendix section) shows that overall the model assumption are held. Table 2 shows the regression output for model 1. Overall we found that the statistically significant predictors for *Classical* are *Instrumentpiano*, *Instrumentstring*, *HarmonyI-V-VI*, and *HarmonyI-V-VI:Voicepar3rd*. The coefficients can be interpreted as follows:

1. Intercept: For stimulus performed in guitar, with harmonic motion I-IV-V, voice leading contrary, the rating provide for as being Classical is 4.25. There are variations based on each test subject on this impact, and the variance of such variation is 2.54

2. Instrumentpiano: For stimulus containing piano instead of guitar, in general the rating of Classical-sounding is expected to increase by 1.37. There are variations based on each test subject on this impact, and the variance of such variation is 1.63.

3. Instrumentstring: For stimulus containing string instead of guitar, in general the rating of Classical-sounding is expected to increase by 3.13. There are variations based on each test subject on this impact and the variance of such variation is 3.51.

4. HarmonyI-V-IV: For stimulus containing harmony I-V-IV instead of I-IV-V, in general the rating of Classical-sounding is expected to be the same as per I-IV-V.

5. HarmonyI-V-VI: For stimulus containing harmony I-V-VI instead of I-IV-V, in general the rating of Classical-sounding is expected to increase by 1.14. There are variations based on each test subject on this impact and the variance of such variation is 1.53.

6. HarmonyI-V-VI: For stimulus containing harmony I-V-VI instead of I-IV-V, in general the rating of Classical-sounding is expected to be the same as per I-IV-V.

7. Voicepar3rd: For stimulus containing voice leading par3rd instead of contrary, in general the rating of Classical-sounding is to be the same .

8. Voicepar5th: For stimulus containing voice leading par5th instead of contrary, in general the rating of Classical-sounding is expected to be the same.

9. HarmonyI-V-IV:Voicepar3rd: For stimulus containing harmony I-V-IV instead of I-IV-V and voice leading par3rd instead of contrary, in general the rating of Classical-sounding does not differ.

10. HarmonyI-V-VI:Voicepar3rd: For stimulus containing harmony I-V-VI instead of I-IV-V and voice leading par3rd instead of contrary, in general the rating of Classical-sounding is expected to decrease by 0.68.

11. HarmonyIV-I-V:Voicepar3rd: For stimulus containing harmony IV-I-V instead of I-IV-V and voice leading par3rd instead of contrary, in general the rating of Classical-sounding does not differ.

12. HarmonyI-V-IV:Voicepar5th: For stimulus containing harmony I-V-IV instead of I-IV-V and voice leading par5th instead of contrary, in general the rating of Classical-sounding does not differ.

13. HarmonyI-V-VI:Voicepar5th: For stimulus containing harmony I-V-VI instead of I-IV-V and voice leading par5th instead of contrary, in general the rating of Classical-sounding does not differ.

14. HarmonyIV-I-V:Voicepar5th: For stimulus containing harmony IV-I-V instead of I-IV-V and voice leading par5th instead of contrary, in general the rating of Classical-sounding does not differ.

The residual analysis for Model 2 (see page 30 to 38 in Appendix section) shows that overall the model assumptions are held. Table 3 shows the regression output for model 2. Overall we found that the statistically significant predictors for $Popular$ are $Instrumentpiano$, $Instrumentstring$, $HarmonyI\text{-}V\text{-}VI$, $HarmonyIV\text{-}I\text{-}V$ and $Voicepar5th$. The coefficients can be interpreted as follows:

1. Intercept: For stimulus performed in guitar, with harmonic motion I-IV-V, voice leading contrary, the rating provide for as being Popular is 6.84. There are variations based on each test subject on this impact, and the variance of such variation is 3.04

2. Instrumentpiano: For stimulus containing piano instead of guitar, in general the rating of popular-sounding is expected to decrease by 1.15. There are variations based on each test subject on this impact, and the variance of such variation is 1.72.

3. Instrumentstring: For stimulus containing string instead of guitar, in general the rating of popular-sounding is expected to decrease by 3.02. There are variations based on each test subject on this impact, and the variance of such variation is 2.55.

4. HarmonyI-V-IV: For stimulus containing harmony I-V-IV instead of I-IV-V, in general the rating of popular-sounding does not differ.

5. HarmonyI-V-VI: For stimulus containing harmony I-V-VI instead of I-IV-V, in general the rating of popular-sounding is expected to decrease by 0.25.

6. HarmonyIV-I-V: For stimulus containing harmony IV-I-V instead of I-IV-V, in general the rating of popular-sounding is expected to decrease by 0.25.

7. Voicepar3rd: For stimulus containing voice leading par3rd instead of contrary, in general the rating of popular-sounding does not differ.

8. Voicepar5th: For stimulus containing voice leading par5th instead of contrary, in general the rating of popular-sounding is expected to increase by 0.23.

Table 2: Regression output for Model 1

	Dependent Variable: *Classical*
(Intercept)	4.25^{***}
	(0.22)
Instrumentpiano	1.37^{***}
	(0.17)
Instrumentstring	3.13^{***}
	(0.24)
HarmonyI-V-IV	0.16
	(0.15)
HarmonyI-V-VI	1.14^{***}
	(0.21)
HarmonyIV-I-V	-0.13
	(0.15)
Voicepar3rd	-0.27
	(0.15)
Voicepar5th	-0.24
	(0.15)
HarmonyI-V-IV:Voicepar3rd	-0.37
	(0.22)
HarmonyI-V-VI:Voicepar3rd	-0.68^{**}
	(0.22)
HarmonyIV-I-V:Voicepar3rd	0.49^{*}
	(0.22)
HarmonyI-V-IV:Voicepar5th	-0.19
	(0.22)
HarmonyI-V-VI:Voicepar5th	-0.43^{*}
	(0.22)
HarmonyIV-I-V:Voicepar5th	0.08
	(0.22)
AIC	9937.27
BIC	10146.84
Log Likelihood	-4932.64
Num. obs.	2493
Num. groups: Subject	70
Var: Subject (Intercept)	2.54
Var: Subject Instrumentpiano	1.63
Var: Subject Instrumentstring	3.51
Var: Subject HarmonyI-V-IV	0.03
Var: Subject HarmonyI-V-VI	1.53
Var: Subject HarmonyIV-I-V	0.00
Var: Residual	2.42

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Regression output for Model 2

	Dependent Variable: *Popular*
(Intercept)	6.84^{***}
	(0.21)
Instrumentpiano	-1.15^{***}
	(0.23)
Instrumentstring	-3.02^{***}
	(0.27)
HarmonyI-V-IV	0.02
	(0.13)
HarmonyI-V-VI	-0.25^{*}
	(0.13)
HarmonyIV-I-V	-0.25^{*}
	(0.13)
Voicepar3rd	0.20
	(0.11)
Voicepar5th	0.23^{*}
	(0.11)
AIC	6358.46
BIC	6438.56
Log Likelihood	-3164.23
Num. obs.	1541
Num. groups: Subject	43
Var: Subject (Intercept)	1.27
Var: Subject Instrumentpiano	1.72
Var: Subject Instrumentstring	2.55
Var: Residual	3.04

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

To address the question of which experiment factor has the strongest impact to people's rating on music genre, we shall refer to the coefficients in our results. From both Model 1 and 2, we noticed that the regression coefficients for $Instrumentpiano$ and $Instrumentstring$ are much larger than the coefficients for the $Harmony$ and $Voice$ levels, suggesting that indeed the instrument used has the strongest impact for listener's music genre recognition. The p-value for the $Instrument$ variables in both Model 1 and 2 are also very small, suggesting the $Instrument$ is a strongly statistical significant predictor for the music genre rating.

We are interested in learning whether the I-V-VI harmonic motion has the strongest association with listener's rating of music pieces as Classical. As per Table 2 shows, we discovered that *HarmonyI-V-VI* is indeed the only statistical significant predictor among all $Harmony$ levels. Also by comparing the coefficient values, we notice that indeed the coefficient for *HarmonyI-V-VI* is the largest among all levels. This supports our hypothesis that the I-V-VI harmony has the strongest association with listener's rating as Classical. We also noticed that the interaction terms *HarmonyI-V-VI:Voicepar3rd* and *HarmonyI-V-VI:Voicepar5th* are statistically significant with negative coefficients. This suggest that when fixing harmonic progression as I-V-VI, if the voice leading is not in contrary motion, the positive association between *HarmonyI-V-VI* and *Classical* will be smaller. Additionally, we have tested whether there exist interactions between $KnowRob$ and $Harmony$ and $KnowAxis$ and $Harmony$ that should be added in our model. The partial F-test result and AIC/BIC values suggested that such interaction terms do not need to be added in our model. Our procedure for performing such checks can be found on page 17 of the Appendix.

To understand whether the contrary motion have the strongest association with listener's rating as Classical among all voice leading types, we have updated the fit for Model 1 in R language so that all levels of $Voice$ is used as predictors and no reference category is taken. The steps can be found on page 16 to 17 of the Appendix section. Table 4 shows the regression result for Model 1 with such updates. We notice that the coefficient for $Voicecontrary$ is the largest among the three levels of $Voice$, suggesting that indeed having the contrary motion have largest positive association with people's rating for a piece as more classical-sounding. However, it as also important to note that $Voice$ is not a statistically significant predictors in our model as per Table 2 shows. Referring to the interaction terms in the model, we also noticed that only when $Voice$ is interacting with *HarmonyI-V-VI*, which we have discovered as an important predictor for Classical music rating, gives statistically significant results. This again suggests that the variable $Voice$ is not an important predictor for $Classical$.

We have discovered that it is needed to include random effects regarding the test subjects in both Model 1 and Model 2. For Model 1, we have included random intercept and random slopes for $Harmony$ and $Instrument$. Referring to Table 2, which is the regression result for Model 1 on $Classcial$, the intercept variance and the variance of the random slope for $Instrumentstring$ is larger than the residual variance, suggesting adding the random intercept and random slope on $Instrument$ is necessary for the model. Although the variance of the random slope of $Harmony$ is smaller than the residual variance, as table 5 shows that the AIC and BIC values have decreased by adding the random effects in the models, we will keep the random slope on $Harmony$ in our model for $Classical$. For Model 2, which has *Popular* as the outcome variable, table 3 shows that the variance for random intercept and random slope on $Instrument$ as smaller than the residual variance. However, as per Table 6 shows that the AIC and BIC values have significantly decreased by adding the random effects in the models, we will keep the two random effects in our model for $Popular$.

Table 4: Regression output for Model 1 with all levels of Voice used as variables

	Dependent Variable: *Classical*
Voicecontrary	4.25^{***}
	(0.22)
Voicepar3rd	3.98^{***}
	(0.22)
Voicepar5th	4.02^{***}
	(0.22)
HarmonyI-V-IV	0.16
	(0.15)
HarmonyI-V-VI	1.14^{***}
	(0.21)
HarmonyIV-I-V	-0.13
	(0.15)
Instrumentpiano	1.37^{***}
	(0.17)
Instrumentstring	3.13^{***}
	(0.24)
Voicepar3rd:HarmonyI-V-IV	-0.36
	(0.22)
Voicepar5th:HarmonyI-V-IV	-0.19
	(0.22)
Voicepar3rd:HarmonyI-V-VI	-0.68^{**}
	(0.22)
Voicepar5th:HarmonyI-V-VI	-0.43^{*}
	(0.22)
Voicepar3rd:HarmonyIV-I-V	0.49^{*}
	(0.22)
Voicepar5th:HarmonyIV-I-V	0.08
	(0.22)
AIC	9936.98
BIC	10146.54
Log Likelihood	-4932.49
Num. obs.	2493
Num. groups: Subject	70
Var: Subject (Intercept)	2.54
Var: Subject Instrumentpiano	1.64
Var: Subject Instrumentstring	3.51
Var: Subject HarmonyI-V-IV	0.04
Var: Subject HarmonyI-V-VI	1.59
Var: Subject HarmonyIV-I-V	0.01
Var: Residual	2.42

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: AIC, BIC values for Model 1 with no random effect, random intercept added, and random slopes on $Instrument$ and $Harmony$ added

	Df	AIC	BIC
fixed effect only	16	10458.10	10551.24
random intercept added	21	10084.96	10207.21
random slopes added	36	9937.27	10146.84

Table 6: AIC, BIC values for Model 2 with no random effect, random intercept added, and random slope on $Instrument$ added

	Df	AIC	BIC
fixed effect only	9	6910.50	6958.56
random intercept added	10	6512.96	6566.36
random slopes added	15	6358.46	6438.56

3.3 Covariates for Genre Recognition

To identify whether there are other important covariates associated with one's recognition of music genre as Classical or Popular, we have performed further variable selection via stepAIC() and fitLMER.fnc() functions in R. The data set used for finding the covariate have been cleaned with all rows with missing values omitted.

To find the covariates for $Classical$, we first used stepAIC() to find a list of candidate fixed effects. Then, we have added those fixed effects into the model, and used the fitLMER.fnc() to test whether those fixed effects were agreed by the function. Finally, we have tested whether for those fixed effects, a random slope regarding the test subject is needed in the model. A detailed procedure of the selection process can be found on page 38 to 43 in the Appendix section. For $Classical$, we have added $PianoPlay$ as the additional covariate, and the purposed model is as follows:

$$\begin{aligned} Classical_i = \alpha_{0j[i]} + \alpha_{1j[i]} Instrument + \alpha_{2j[i]} Harmony + \beta_3 Voice + \\ \beta_4 Harmony * Voice + \beta_5 PianoPlay + \epsilon_i, \epsilon_i \sim_{iid} N(0, \sigma^2) \\ \alpha_{0j} = \beta_0 + \eta_{0j}, \eta_{0j} \sim_{iid}, N(0, \tau_0^2) \\ \alpha_{1j} = \beta_1 + \eta_{1j}, \eta_{1j} \sim_{iid}, N(0, \tau_1^2) \\ \alpha_{2j} = \beta_2 + \eta_{2j}, \eta_{2j} \sim_{iid}, N(0, \tau_2^2) \end{aligned} \tag{3}$$

The residual analysis for Model 3 (see page 44 to 46 in Appendix section) shows that overall the model assumptions are held. The regression output is presented in Table 7. The newly added covariate $PianoPlay$ appeared to be an important predictor of $Classical$, as the coefficients for $PianoPlay4$ and $PianoPlay5$ are statistically significant. A random slope of $PianoPlay$ regarding $Subject$ was not added in Model 3 as fitLMER.fnc() function suggested that such random slope is not needed in the model. The coefficients of $PianoPlay$ can be interpreted as follows:

1. PianoPlay1: For people identify themselves as scale = 1 of how often they play piano instead of scale = 0, in general the rating of Classcial-sounding does not differ.
2. PianoPlay4: For people identify themselves as scale = 4 of how often they play piano instead of scale = 0, in general the rating of Classcial-sounding is expected to increase by 1.99.

3. PianoPlay5: For people identify themselves as scale = 5 of how often they play piano instead of scale = 0, in general the rating of Classcial-sounding is expected to increase by 1.09.

For *Popular*, we have also tested whether additional fixed effect covariates should be added to Model 2 with the same set of procedures mentioned above. Although fitLMER.fnc() with three different methods (AIC, t-test and log ratio test) have all suggested to add *PachListen* as fixed effect to Model 2, we discovered that the lmer() function would fail to converge after adding *PachListen*. Hence, we have decided to not add any additional covariates to Model 2. A detailed procedure of this process can be viewed in page 46 to 48 of the Appendix section.

3.4 Impact of Self-identification as Musician on Genre Recognition

To analyze whether one's self-identification as musician is associated with the recognition of music genre as Classical or Popular, we have first investigated the appropriate cut-off to dichotomize the variable *Selfdeclare*. Based on histograms, we found that setting the cut-off at 3 would yield a roughly 50-50 split of all responses. A detailed steps for such process can be found on page 48 to 49 of the Appendix section.

To check whether *Selfdeclare* is associated directly with the music genre recognition, we have first tested whether the dichotomized *Selfdeclare* variable should be added as fixed effect to the best model for *Classical* (Model 3) and *Popular* (Model 2). The partial F-test results per Table 8 and 9 show that for modeling *Classical* and *Popular*, the dichotomized *Selfdeclare* term is not needed.

As we are interested in learning whether if self-identification as musician impacts how people use *Instrument*, *Harmony*, *Voice*, and *PianoPlay* to generate the rating of music genre, we have further tested whether interaction between *Selfdeclare* and the existing fixed effects should be added to Model 2 and 3 via partial F-tests. The F-tests results suggest that for *Classical*, adding such interaction terms is not necessary. For *Popular*, we discovered that the interaction between the dichotomized *Selfdeclare* and *Harmony* should be added in the model, as the partial F-test result (Table 9) shows a p-value = 0.000983, suggesting that the model with interaction added is preferred. Hence, we propose that the best model for *Popular* as

$$\begin{aligned} Popular_i = \alpha_{0j[i]} + \alpha_{1j[i]} Instrument + \beta_2 Harmony + \beta_3 Voice + \\ \beta_4 Selfdeclare + \beta_5 Harmony * Selfdeclare + \epsilon_i, \epsilon_i \sim_{iid} N(0, \sigma^2) \\ \alpha_{0j} = \beta_0 + \eta_{0j}, \eta_{0j} \sim_{iid} N(0, \tau_0^2) \\ \alpha_{1j} = \beta_1 + \eta_{1j}, \eta_{1j} \sim_{iid} N(0, \tau_1^2) \end{aligned} \tag{4}$$

The residual analysis for Model 4 (see page 62 to 65 in Appendix section) shows that overall the model assumptions are held. Table 11 shows the regression output for Model 4. Although the *Harmony* and *Selfdeclare* variables themselves are not statistically significant predictors, the interaction term *HarmonyI-V-VI:Selfdeclare* appears to be an important predictor as *HarmonyI-V-VI:Selfdeclare* is strongly statistically significant in our regression output. We noticed that after adding the interaction between *Harmony* and *Selfdeclare*, the impact of harmonic motion I-V-VI on the rating for popular genre as switched from positive sign to negative sign, suggesting the strong relationship between the two variables. The coefficient of the interaction terms can be interpreted as follows:

Table 7: Regression output for Model 3

	Dependent Variable: *Classical*
(Intercept)	3.37^{***}
	(0.27)
Instrumentpiano	1.65^{***}
	(0.23)
Instrumentstring	3.59^{***}
	(0.31)
HarmonyI-V-IV	0.21
	(0.20)
HarmonyI-V-VI	1.27^{***}
	(0.28)
HarmonyIV-I-V	-0.30
	(0.20)
Voicepar3rd	-0.31
	(0.19)
Voicepar5th	-0.20
	(0.19)
PianoPlay1	0.61
	(0.36)
PianoPlay4	1.99^{**}
	(0.62)
PianoPlay5	1.09^{*}
	(0.50)
HarmonyI-V-IV:Voicepar3rd	-0.43
	(0.27)
HarmonyI-V-VI:Voicepar3rd	-0.71^{*}
	(0.28)
HarmonyIV-I-V:Voicepar3rd	0.75^{**}
	(0.27)
HarmonyI-V-IV:Voicepar5th	-0.21
	(0.28)
HarmonyI-V-VI:Voicepar5th	-0.52
	(0.28)
HarmonyIV-I-V:Voicepar5th	0.34
	(0.27)
AIC	6184.69
BIC	6392.96
Log Likelihood	-3053.34
Num. obs.	1541
Num. groups: Subject	43
Var: Subject (Intercept)	1.33
Var: Subject Instrumentpiano	1.93
Var: Subject Instrumentstring	3.68
Var: Subject HarmonyI-V-IV	0.14
Var: Subject HarmonyI-V-VI	1.76
Var: Subject HarmonyIV-I-V	0.14
Var: Residual	2.43

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Partial F-test result for Model 3 ($Classical$) on $Selfdeclare$ added as fixed effect

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 3	39	6184.69	6392.96	-3053.34	6106.69			
$Selfdeclare$ added	40	6186.75	6400.35	-3053.37	6106.75	0.00	1	1.0000

Table 9: Partial F-test result for Model 2 ($Popular$) on $Selfdeclare * Harmony$ added

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 2	15	6358.46	6438.56	-3164.23	6328.46			
$Selfdeclare$ added	16	6360.46	6445.90	-3164.23	6328.46	0.00	1	0.9860
$Selfdeclare$ and $Selfdeclare * Harmony$ added	19	6350.16	6451.62	-3156.08	6312.16	16.30	3	0.0010

1. HarmonyI-V-IV:Selfdeclare: For people identify themselves as musicians, for music pieces with the harmonic motion I-V-IV instead of I-IV-V the rating of popular-sounding does not differ (as HarmonyI-V-IV is not statistically significant) .

2. HarmonyI-V-VI:Selfdeclare: For people identify themselves as musicians, for music pieces with the harmonic motion I-V-VI instead of I-IV-V the rating of popular-sounding is expected to decrease by 0.70 ($0.13 - 0.83 = -0.70$).

3. HarmonyIV-I-V:Selfdeclare: For people identify themselves as musicians, for music pieces with the harmonic motion IV-I-V instead of I-IV-V the rating of popular-sounding does not differ (as HarmonyIV-I-V is not statistically significant).

We have tested whether such result holds if switching the dichotomization cut-off from 3 to 2 and 4. Our partial F-test results suggested that when dichotomizing $Selfdeclare$ at 2, the interaction term $Harmony * Selfdeclare$ is very close to being statistically significant, whereas dichotomizing $Selfdeclare$ at 4 would again yield $Harmony*Selfdeclare$ as statistically significant predictor with negative coefficient. This suggests that the interaction $Harmony * Selfdeclare$ is quite important to keep in our model for $Popular$.

4 Discussion

To address the question of whether instrument used has a stronger impact on people's recognition of music genre than the harmonic motion and voice leading of a music piece, we have constructed two mixed models with the three factors. Our model result suggests that indeed for both classical and popular genre, instrument has the strongest impact on people's rating on how the music piece sound

Table 10: AIC, BIC values for Model 3 with no random intercept and random intercept regarding $Selfdeclare$ added

	Df	AIC	BIC
no random intercept	39	6184.69	6392.96
random intercept added	40	6187.35	6400.96

Table 11: Regression output for Model 4

	Dependent Variable: *Popular*
(Intercept)	6.75^{***}
	(0.28)
Instrumentpiano	-1.15^{***}
	(0.23)
Instrumentstring	-3.02^{***}
	(0.27)
HarmonyI-V-IV	-0.00
	(0.17)
HarmonyI-V-VI	0.13
	(0.17)
HarmonyIV-I-V	-0.22
	(0.17)
Voicepar3rd	0.20
	(0.11)
Voicepar5th	0.23^{*}
	(0.11)
Selfdeclare	0.20
	(0.38)
HarmonyI-V-IV:Selfdeclare	0.05
	(0.25)
HarmonyI-V-VI:Selfdeclare	-0.83^{***}
	(0.25)
HarmonyIV-I-V:Selfdeclare	-0.07
	(0.25)
AIC	6350.16
BIC	6451.62
Log Likelihood	-3156.08
Num. obs.	1541
Num. groups: Subject	43
Var: Subject (Intercept)	1.28
Var: Subject Instrumentpiano	1.73
Var: Subject Instrumentstring	2.56

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

as those genres than the two other factors. Such phenomenon is as expected. As strings are among the most frequently used instruments for orchestras that perform classical music, and many famous classical pieces were specifically written for piano (such as the Moonlight Sonata by Beethoven), it comes no surprise that people tend to associate music performed in these two instruments as more classical-sounding. Similarly, as guitar is a relatively modern instrument and is frequently used in more popular music genres such as Rock, people tend to associate music with guitar as more popular-sounding. Our finding of the importance of instrument on music genre recognition supports our intuitive understanding of the subject.

We have also investigated whether the harmonic motion I-V-Vi have the strongest association with Classical ratings among all harmonic motion types, and whether such association also depends on whether the listener is familiar with the Pachelbel rants by comedian Robert Paravonian, or the comedy bits by Axis of Evil on the Pachelbel's chords. As variable I-V-Vi was strongly significant and returns the largest coefficient among all harmonic motion levels in our model result, it suggests that indeed I-V-Vi have the strongest positive association with Classical ratings. We have also discovered that there are no interaction terms of the variable $Harmony$ with $KnowRob$ and $KnowAxis$ in our model for $Classical$, suggesting that the association between I-V-Vi and classical rating does not depend on the listener's familiarity with Pachelbel Rant or comedy bits. This observation makes intuitive sense. The harmonic progression I-V-Vi is known as "Pachelbel's progression" as it was featured in Pachelbel's Canon in D, which a widely famous and representative piece of the Baroque style music. Hence, it is natural that people automatically associate pieces with such chord progression with the classical genre regardless whether they have seen comedy bits or popular commentary on the harmonic progression. In general, our finding that harmonic progression I-V-Vi is a strong indicator for one's recognition of music as the classical genre satisfies our understanding on the popularity of this specific harmonic progression.

Additionally, we have discovered that there exists an interaction for the harmonic motion and voice leading for predicting the Classical genre rating, while such interaction is not needed for Popular music ratings. Our model suggests that when the voice leading is not in contrary motion, the positive association between harmonic progression I-V-VI and Classical rating would become weaker. This observation should come with no surprise. As the practice of using parallel fifth was strongly discouraged for composing during the common practice period (Wikipedia.org, 2019), in general there were much less classical pieces with parallel fifth or third as opposed to contrary motion. Hence even when the piece is presented with harmony I-V-VI, the fact that it is presented in parallel fifth or third would decrease people's recognition of the piece as Classical genre. Hence, we should take close note to the interaction between harmonic motion and voice leading when analyzing the individual effect of of these two factors on rating as Classical.

We have investigated whether the contrary motion has the strongest association with one's rating for music as Classical among all types of voice leading types. Our model suggests that although contrary motion does have the largest coefficient among all levels of $Voice$ in our regression model, overall voice leading itself does not appear to be a statistically significant predictor for people's rating of Classical genre. By examining the model, we noticed that although $Voice$ is not statistically significant, the interaction term $Voice * Harmony$ was strongly significant for the model predicting $Classical$. Therefore, we should pay close attention to not only interpreting the stand-alone effect of voice leading on the Classical rating, but rather to conduct analysis on how the specific voice leading types affect the impact of different harmonic motions on the Classical rating. Overall, although the coefficient for contrary motion was the largest among all $Voice$ levels for modeling $Classical$, as the variable itself is not statistically significant, we should take this result with a pinch of salt.

For both models predicting Classical and Popular, we have discovered random effects are needed in the model to achieve better fits. We found that for music pieces performed in guitar, with harmonic motion I-IV-V, and in contrary motion, there exists a variation on the rating as Classical and Population provided by each individual. Further, we have discovered there exists a per-subject variation on how people generate the genre rating as Classical when the music piece is presented in string (as the random effect variance was larger than residual variance). Interestingly, we have received contradictory findings regarding whether there also exists per-subject bias on how people provide Classical ratings based on the harmonic motion, and how people provide Popular ratings based on the instrument. In our models, the variance of these random effects were smaller than the residual variance, suggesting these random effects is perhaps not as important. However, the AIC and BIC values suggested that adding these random effects in the model improves the model quality. Due to these contradictory statistical findings, it is unclear whether the per-subject bias exists regarding harmony in the Classical model, and instrument in the Popular model. A future research direction is to further investigate whether such per-subject biases do impact how people distinguish Classical music based on harmonic motion, and for Popular music based on instrument used.

We have also investigated whether there are more important factors besides the three main experiment factors that are associated with the Classical or Popular ratings. For the Popular rating, we found that having the three main factors (instrument, harmonic motion, voice leading) as predictors is sufficient and no extra variable is needed. For the Classical genre, our results suggest that one's familiarity with piano appears to have a positive relationship with their recognition of music as Classical. For people that are familiar with the instrument, they are more likely to rate music pieces in general as more Classical-sounding. This is an interesting observation, as it indicates that people who are skilled in playing piano are in general more inclined to categorize all music as Classical genre, regardless of other factors such as instrument used of the music piece. For future research, more investigation can be conducted to analyze if there are specific reasons such as psychological factors that causes the positive association between piano skills and rating for music as Classical.

As we are interested in whether one self-identifies as musician changes the dynamic of how instrument, harmonic motion, and voice leading interact with the rating provided for music genre, we have checked the significance of interactions between a dichotomized $Selfdeclare$ variable with the three main experiment factors for both $Classical$ and $Popular$. Our finding suggests that for the Classical genre, whether one self-identifies as musician does not change the relationship between the genre rating and experiment factors. However, there exists a significant interaction between harmonic progression I-V-VI with $Selfdeclare$ on the rating given for music as popular genre, and such finding is quite robust regardless of the specific cut-off values used to define musicians and non-musicians. For those that consider themselves as musicians, they tend to associate harmonic progression I-V-VI with less popular-sounding, whereas those do not consider themselves as musicians would find no association between such harmonic progression and the popular rating. Such phenomenon could be due to the musicians having more knowledge of music theory and are aware of the I-V-VI's appearance in the famous Pachelbel's Canon, hence not categorizing songs with I-V-VI as the Popular genre. Overall, we have discovered that for people self-identifies as musician, the dynamic between how they associate harmonic motion I-V-Vi differs from those who does not.

There are also shortcomings in our model. The first issue is that our approach of data cleaning could have potentially caused the model fitted not reflecting important patterns in the original data set, or have introduced bias to the fitted model. As mentioned in Section 3.1, we have performed data cleaning by omitting all entries with missing values in our data, and used the cleaned data set to develop models that focused on finding additional covariates in the data set that impacts the Classical and Popular rating. By omitting the observations with missing values, we have reduced the data set from 2520 to 1506 entries. Although the reduced data set still contains reasonable amount of observations to work with, it turned out that the deletion of the missing value rows have removed some interesting patterns within the data set. For example, for the variable $PianoPlay$, the observations in the uncleaned data set include value of 0, 1, 2, 4, 5. After the cleaning, all observations with a response of 2 were eliminated. As in Section 3.3 we have discovered that $PianoPlay$ was indeed an important covariate of $Classical$, our model generated with $PianoPlay$ as predictor fails to reflect the dynamic between one's piano skill and the rating provided for when $PianoPlay = 2$. For future reference, imputation methods such as multiple imputation or regression imputation should be performed on the data set to avoid omitting observations to avoid introducing bias to the model, and develop model that accounts for more information regarding the original data set.

The second problem is that the definition used for distinguishing musicians and non-musicians in our model is rather ambiguous and not robust. We have dichotomized the variable $Selfdeclare$, which is a self-reported rating on the degree of people consider themselves as musicians. As the definition of "musician" could vary greatly between person, this strongly subjective variable might not be very consistent, hence leading our finding of association between musicians and the Popular rating as not very robust. An approach to improve this is to define musician and non-musicians using more objective, quantitative variables. For example, we can use the OMSI scores, which has a standardized interpretation of the scoring levels. The respondents with a score greater than 500 are classified as "more musically sophisticated" and those with a score less than 500 as "less musically sophisticated" (Marcs-survey.uws.edu.au, 2019). Other potential variables such as number of concerts performed, or years of music training received can also be used to classify musicians and non-musician. By using more objective variables to construct the musician variable, the model constructed will be capable of showing a clearer, more scientific-robust picture on how the musicians uses information such as instrument differently to distinguish music genre.

References

En.wikipedia.org. (2019). Music Genre (United States). [online] Available at: https://en.wikipedia.org/wiki/Music

En.wikipedia.org. (2019). Consecutive Fifths (United States). [online] Available at: https://en.wikipedia.org/wiki/

Lena, J., Peterson, R. (2008). Classification as Culture: Types and Trajectories of Music Genres. *American Sociological Review*, 73(5), 697-718. Retrieved from www.jstor.org/stable/25472554

Marcs-survey.uws.edu.au. (2019). OMSI Online. [online] Available at: (http://marcs-survey.uws.edu.au/OMSI/om

Oramas, S., Barbieri, F., Nieto, O. and Serra, X., 2018. Multimodal Deep Learning for Music Genre Classification. *Transactions of the International Society for Music Information Retrieval*, 1(1), 4–21.

Tech Appendix project 3

Yichun Shi

12/3/2019

EDA and data cleaning

```r
library(skimr)
```

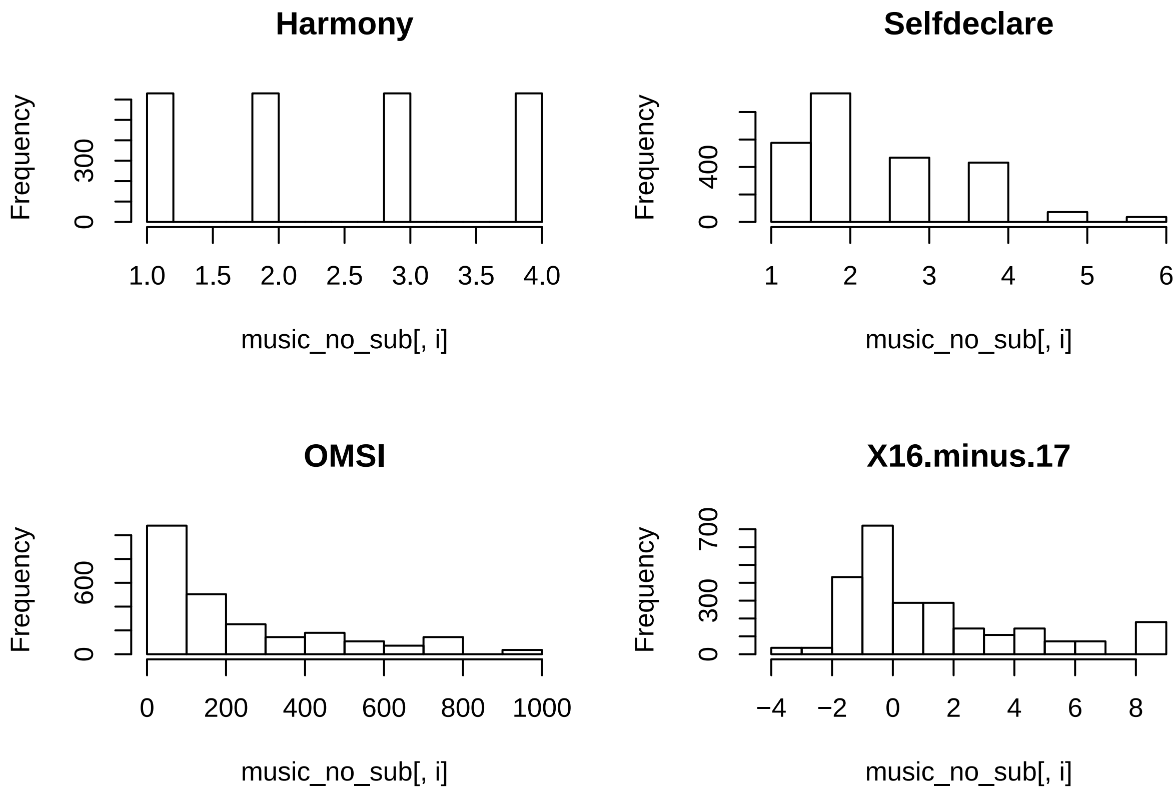
```
##
## Attaching package: 'skimr'

## The following object is masked from 'package:stats':
##
##     filter
```

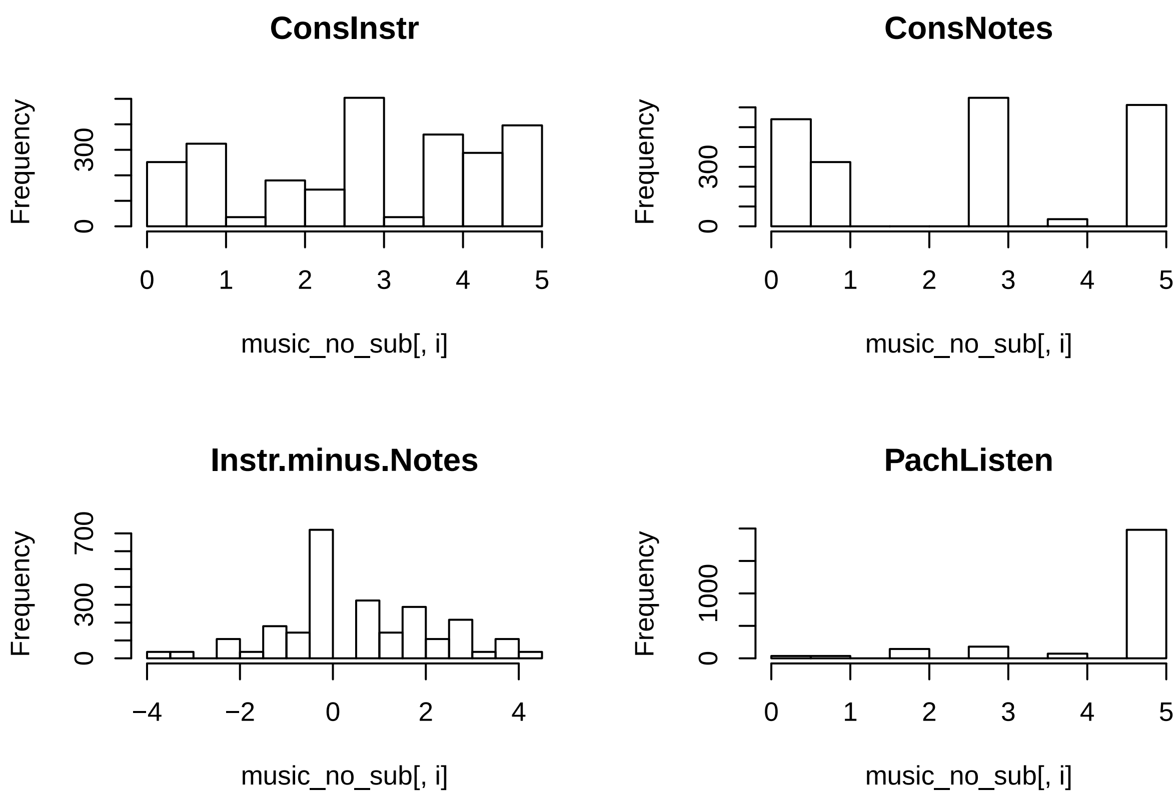
```r
skim(music)
```

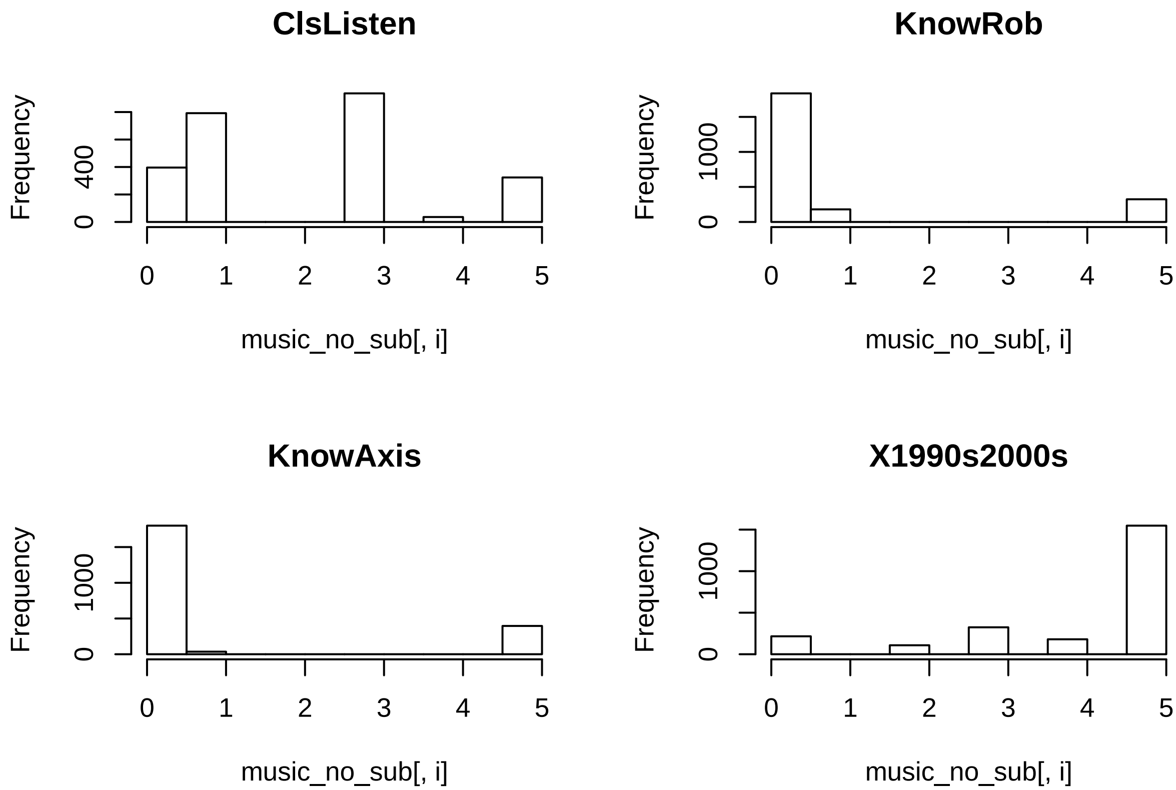
```
## Skim summary statistics
##  n obs: 2520
##  n variables: 28
##
## -- Variable type:factor -------------------------------------------------------------------------
##    variable missing complete    n n_unique
##     first12       0     2520 2520        3
##     Harmony       0     2520 2520        4
##  Instrument       0     2520 2520        3
##     Subject       0     2520 2520       70
##       Voice       0     2520 2520        3
##                              top_counts ordered
##    str: 1080, gui: 720, pia: 720, NA: 0   FALSE
##  I-I: 630, I-V: 630, I-V: 630, IV-: 630   FALSE
##     gui: 840, pia: 840, str: 840, NA: 0   FALSE
##         15: 36, 16: 36, 17: 36, 18b: 36   FALSE
##     con: 840, par: 840, par: 840, NA: 0   FALSE
##
## -- Variable type:integer ------------------------------------------------------------------------
##                     variable missing complete    n    mean     sd p0
##                     APTheory     216     2304 2520    0.23   0.42  0
##                    ClsListen      36     2484 2520    2.16   1.59  0
##                 CollegeMusic     108     2412 2520    0.79   0.41  0
##                    Composing      72     2448 2520    1      1.46  0
##                    ConsNotes     360     2160 2520    2.53   1.95  0
##                   GuitarPlay       0     2520 2520    0.69   1.48  0
##                     KnowAxis     288     2232 2520    0.9    1.91  0
##                      KnowRob     180     2340 2520    0.77   1.72  0
##                      NoClass     288     2232 2520    0.92   1.5   0
##                         OMSI       0     2520 2520  225.93 231.32 11
##                   PachListen      72     2448 2520    4.51   1.1   0
```

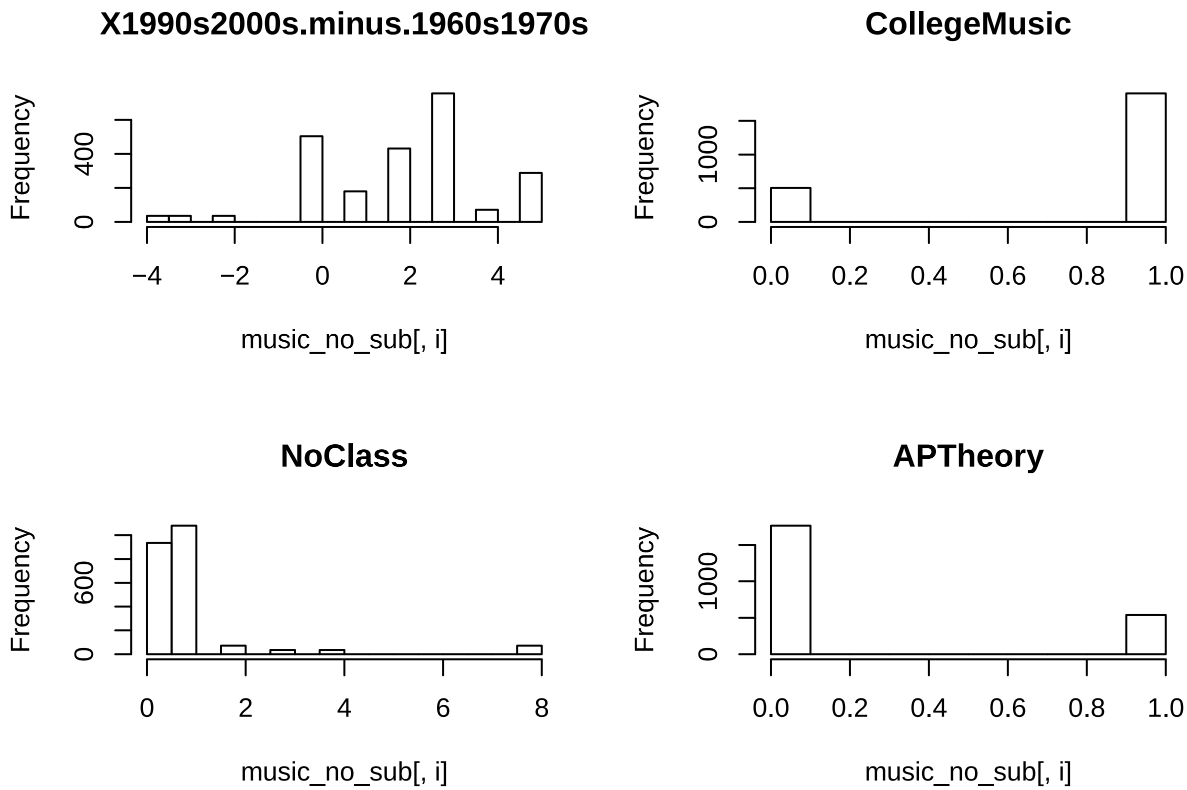
```
##                   PianoPlay       0     2520 2520    1.09   1.72  0
##                  Selfdeclare       0     2520 2520    2.44   1.18  1
##                            X       0     2520 2520 1260.5  727.61  1
##                  X1990s2000s     144     2376 2520    4.06   1.56  0
##  X1990s2000s.minus.1960s1970s     180     2340 2520    2.02   1.92 -4
##                    X1stInstr    1512     1008 2520    2.79   1.59  1
##                    X2ndInstr    2196      324 2520    1.56   1.17  0
##     p25    p50     p75 p100     hist
##    0       0      0       1 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2582>
##    1       3      3       5 <U+2583><U+2587><U+2581><U+2581><U+2587><U+2581><U+2581><U+2583>
##    1       1      1       1 <U+2582><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2587>
##    0       0      2       5 <U+2587><U+2582><U+2581><U+2582><U+2581><U+2581><U+2582><U+2581>
##    0.75    3      5       5 <U+2587><U+2583><U+2581><U+2581><U+2587><U+2581><U+2581><U+2587>
##    0       0      1       5 <U+2587><U+2582><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##    0       0      0       5 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2582>
##    0       0      0       5 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2582>
##    0       1      1       8 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##   49     145.5  323     970 <U+2587><U+2585><U+2582><U+2581><U+2581><U+2581><U+2581><U+2581>
##    5       5      5       5 <U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2587>
##    0       0      1       5 <U+2587><U+2583><U+2581><U+2581><U+2581><U+2581><U+2581><U+2582>
##    2       2      3       6 <U+2585><U+2587><U+2581><U+2583><U+2583><U+2581><U+2581><U+2581>
##  630.75 1260.5 1890.25 2520 <U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587>
##    3       5      5       5 <U+2581><U+2581><U+2581><U+2581><U+2582><U+2581><U+2581><U+2587>
##    0       2      3       5 <U+2581><U+2581><U+2581><U+2585><U+2582><U+2585><U+2587><U+2583>
##    1       3.5    4       5 <U+2587><U+2582><U+2581><U+2581><U+2581><U+2587><U+2581><U+2583>
##    1       1      2       4 <U+2582><U+2587><U+2581><U+2582><U+2581><U+2582><U+2581><U+2582>
##
## -- Variable type:numeric --------------------------------------------------------------------
##           variable missing complete    n mean   sd p0  p25  p50  p75  p100
##          Classical      27     2493 2520 5.78 2.66  0 4    6    8    19
##          ConsInstr       0     2520 2520 2.86 1.58  0 1.67 3    4.33  5
##  Instr.minus.Notes       0     2520 2520 0.69 1.69 -4 0    0.34 2     4.33
##            Popular      27     2493 2520 5.38 2.5   0 4    5    7    19
##       X16.minus.17       0     2520 2520 1.72 2.99 -4 0    1    3     9
##      hist
##  <U+2583><U+2585><U+2587><U+2585><U+2582><U+2581><U+2581><U+2581>
##  <U+2583><U+2585><U+2583><U+2583><U+2587><U+2585><U+2586><U+2586>
##  <U+2581><U+2581><U+2582><U+2587><U+2583><U+2583><U+2583><U+2582>
##  <U+2583><U+2583><U+2587><U+2583><U+2581><U+2581><U+2581><U+2581>
##  <U+2581><U+2585><U+2587><U+2586><U+2583><U+2582><U+2582><U+2582>
```

Plotting histograms for variables, notice that OMSI has very large range and is highly skewed

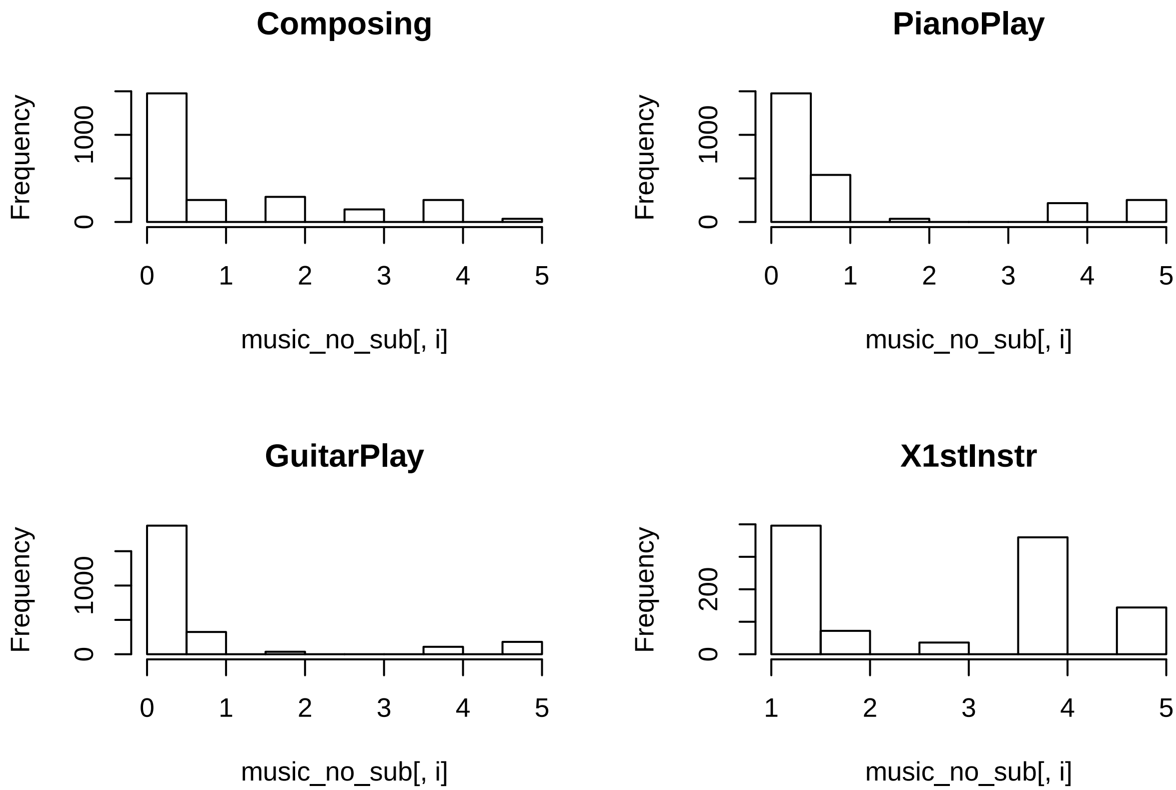

Harmony
Frequency
0
300
1.0
1.5
2.0
2.5
3.0
3.5
4.0
music_no_sub[, i]
Selfdeclare
Frequency
0
400
1
2
3
4
5
6
music_no_sub[, i]
OMSI
Frequency
0
600
0
200
400
600
800
1000
music_no_sub[, i]
X16.minus.17
Frequency
0
300
700
−4
−2
0
2
4
6
8
music_no_sub[, i]

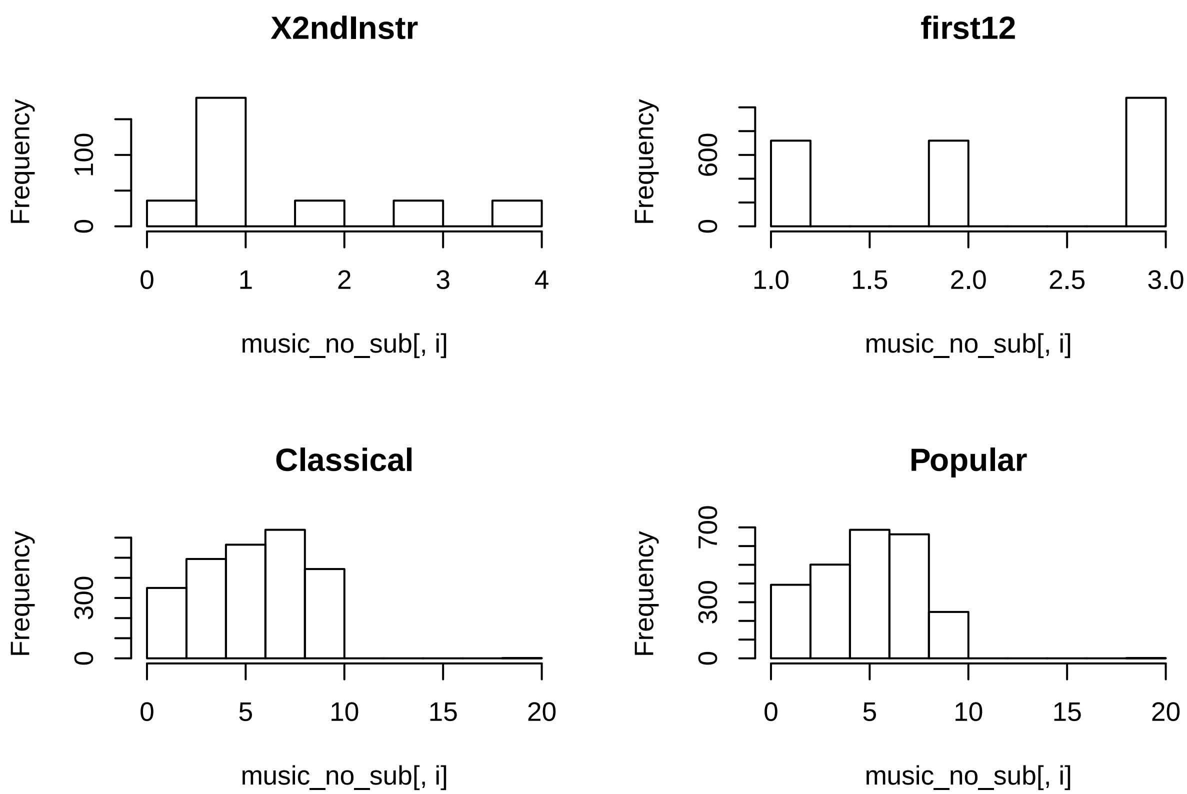

ConsInstr
Frequency
0
300
0
1
2
3
4
5
music_no_sub[, i]
ConsNotes
Frequency
0
300
0
1
2
3
4
5
music_no_sub[, i]
Instr.minus.Notes
Frequency
0
300
700
−4
−2
0
2
4
music_no_sub[, i]
PachListen
Frequency
0
1000
0
1
2
3
4
5
music_no_sub[, i]


ClsListen
KnowRob
KnowAxis
X1990s2000s
Frequency
0
400
1000
0 1 2 3 4 5
music_no_sub[, i]



X1990s2000s.minus.1960s1970s
Frequency
0
400
−4
−2
0
2
4
music_no_sub[, i]
CollegeMusic
Frequency
0
1000
0.0
0.2
0.4
0.6
0.8
1.0
music_no_sub[, i]
NoClass
Frequency
0
600
0
2
4
6
8
music_no_sub[, i]
APTheory
Frequency
0
1000
0.0
0.2
0.4
0.6
0.8
1.0
music_no_sub[, i]


Composing
PianoPlay
GuitarPlay
X1stInstr
Frequency
music_no_sub[, i]
0
1000
200
0
1
2
3
4
5



Data cleaning, the music.1 data set will be used for fitting models regarding the three main experiment factors

```
unique(music$ConsInstr)
```

```
##  [1] 4.33 2.33 1.00 3.67 3.00 5.00 0.00 1.67 0.67 2.67 4.00 3.33 2.00 1.33
```

```
music.1 <- music %>% mutate(ConsInstr = round(ConsInstr))
music.1 <- music.1 %>% filter(is.na(Classical) == F)
```

Preparing new data set music.sub by omitting rows with missing values, log transform OMSI and convert all ordinal and dummary variables to factors. This data set will only be used for finding additional covariates

```
music.sub <- music.1[,c(2:23,27,28)]
music.sub <- na.omit(music.sub)
music.sub <- music.sub %>% mutate(log_OMSI = log(OMSI))

music.sub$Selfdeclare <- as.factor(music.sub$Selfdeclare)
music.sub$PachListen <- as.factor(music.sub$PachListen)
music.sub$ConsInstr<- as.factor(music.sub$ConsInstr)
music.sub$ConsNotes<- as.factor(music.sub$ConsNotes)
music.sub$X1990s2000s<- as.factor(music.sub$X1990s2000s)
music.sub$ClsListen <- as.factor(music.sub$ClsListen)
music.sub$CollegeMusic <- as.factor(music.sub$ CollegeMusic)
music.sub$KnowRob<- as.factor(music.sub$KnowRob)
```

```r
music.sub$KnowAxis<- as.factor(music.sub$KnowAxis)
music.sub$PianoPlay<- as.factor(music.sub$PianoPlay)
music.sub$GuitarPlay<- as.factor(music.sub$GuitarPlay)
music.sub$APTheory<- as.factor(music.sub$APTheory)
```

Fitting Classical with Instrument, Harmony, Voice

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony + Voice + Harmony:Voice,
##     data = music.1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8570 -1.7245  0.0228  1.6715 11.7530
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               4.24840    0.17127  24.806  < 2e-16 ***
## Instrumentpiano           1.37398    0.11277  12.184  < 2e-16 ***
## Instrumentstring          3.13256    0.11209  27.947  < 2e-16 ***
## HarmonyI-V-IV             0.15288    0.22506   0.679   0.4970
## HarmonyI-V-VI             1.14044    0.22478   5.074 4.19e-07 ***
## HarmonyIV-I-V            -0.13397    0.22424  -0.597   0.5503
## Voicepar3rd              -0.27452    0.22424  -1.224   0.2210
## Voicepar5th              -0.23263    0.22478  -1.035   0.3008
## HarmonyI-V-IV:Voicepar3rd -0.35922    0.31770  -1.131   0.2583
## HarmonyI-V-VI:Voicepar3rd -0.68392    0.31808  -2.150   0.0316 *
## HarmonyIV-I-V:Voicepar3rd  0.48855    0.31770   1.538   0.1242
## HarmonyI-V-IV:Voicepar5th -0.19140    0.31846  -0.601   0.5479
## HarmonyI-V-VI:Voicepar5th -0.43173    0.31789  -1.358   0.1746
## HarmonyIV-I-V:Voicepar5th  0.06817    0.31751   0.215   0.8300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.292 on 2479 degrees of freedom
## Multiple R-squared:  0.2596, Adjusted R-squared:  0.2558
## F-statistic: 66.88 on 13 and 2479 DF,  p-value: < 2.2e-16
```

lm.clas2 is fitted based on the output from stepAIC, which is lm.clas1.step

```r
lm.clas2 <- lm(Classical ~ Instrument+ Harmony*Voice, data = music.1)
summary(lm.clas2)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony * Voice, data = music.1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8570 -1.7245  0.0228  1.6715 11.7530
##
```

```
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  4.24840    0.17127  24.806  < 2e-16 ***
## Instrumentpiano              1.37398    0.11277  12.184  < 2e-16 ***
## Instrumentstring             3.13256    0.11209  27.947  < 2e-16 ***
## HarmonyI-V-IV                0.15288    0.22506   0.679   0.4970
## HarmonyI-V-VI                1.14044    0.22478   5.074 4.19e-07 ***
## HarmonyIV-I-V               -0.13397    0.22424  -0.597   0.5503
## Voicepar3rd                 -0.27452    0.22424  -1.224   0.2210
## Voicepar5th                 -0.23263    0.22478  -1.035   0.3008
## HarmonyI-V-IV:Voicepar3rd -0.35922    0.31770  -1.131   0.2583
## HarmonyI-V-VI:Voicepar3rd -0.68392    0.31808  -2.150   0.0316 *
## HarmonyIV-I-V:Voicepar3rd  0.48855    0.31770   1.538   0.1242
## HarmonyI-V-IV:Voicepar5th -0.19140    0.31846  -0.601   0.5479
## HarmonyI-V-VI:Voicepar5th -0.43173    0.31789  -1.358   0.1746
## HarmonyIV-I-V:Voicepar5th  0.06817    0.31751   0.215   0.8300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.292 on 2479 degrees of freedom
## Multiple R-squared:  0.2596, Adjusted R-squared:  0.2558
## F-statistic: 66.88 on 13 and 2479 DF,  p-value: < 2.2e-16
```

```r
anova(lm.clas.base, lm.clas2,lm.clas1)
```

```
## Analysis of Variance Table
##
## Model 1: Classical ~ Instrument + Harmony + Voice
## Model 2: Classical ~ Instrument + Harmony * Voice
## Model 3: Classical ~ Instrument * Harmony * Voice
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1   2485 13108
## 2   2479 13026  6    81.280 2.5714 0.01741 *
## 3   2457 12944 22    82.005 0.7075 0.83618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for random intercept: both AIC and BIC dropped significantly after adding random intercept

```r
lmer.clas2 <- lmer(Classical ~ Instrument + Harmony*Voice + (1|Subject), data = music.1, REML=F)
summary(lmer.clas2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ Instrument + Harmony * Voice + (1 | Subject)
##    Data: music.1
##
##      AIC      BIC   logLik deviance df.resid
##  10458.1  10551.2  -5213.1  10426.1     2477
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8924 -0.6212 -0.0165  0.6392  5.6657
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.678    1.295
##  Residual             3.537    1.881
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                  4.25306    0.20907  20.342
## Instrumentpiano              1.37746    0.09261  14.873
## Instrumentstring             3.13086    0.09200  34.030
## HarmonyI-V-IV                0.14892    0.18469   0.806
## HarmonyI-V-VI                1.14100    0.18445   6.186
## HarmonyIV-I-V               -0.13397    0.18398  -0.728
## Voicepar3rd                 -0.28018    0.18400  -1.523
## Voicepar5th                 -0.23618    0.18444  -1.281
## HarmonyI-V-IV:Voicepar3rd -0.34960    0.26072  -1.341
## HarmonyI-V-VI:Voicepar3rd -0.68277    0.26100  -2.616
## HarmonyIV-I-V:Voicepar3rd  0.49026    0.26068   1.881
## HarmonyI-V-IV:Voicepar5th -0.19316    0.26130  -0.739
## HarmonyI-V-VI:Voicepar5th -0.42874    0.26087  -1.644
## HarmonyIV-I-V:Voicepar5th  0.06604    0.26051   0.254
```

```
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
anova(lmer.clas2,lm.clas2)
```

```
## Data: music.1
## Models:
## lm.clas2: Classical ~ Instrument + Harmony * Voice
## lmer.clas2: Classical ~ Instrument + Harmony * Voice + (1 | Subject)
##            Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm.clas2   15 11227 11314 -5598.5    11197
## lmer.clas2 16 10458 10551 -5213.1    10426 770.84      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for random slopes:

We have set the r code chunk as include=FALSE to hide the long output from fitLMER.fnc(). The code ran here is as follows:

clas2.bic<- fitLMER.fnc(lmer.clas2, ran.effects = c("(0 + Instrument|Subject)","(0 + Harmony|Subject)","(0 + Voice|Subject)"),method="BIC") clas2.t<- fitLMER.fnc(lmer.clas2, ran.effects = c("(0 +Instrument|Subject)","(0 + Harmony|Subject)","(0 + Voice|Subject)"),method="t") clas2.llrt<- fitLMER.fnc(lmer.clas2, ran.effects = c("(0 + Instrument|Subject)","(0 + Harmony|Subject)","(0 + Voice|Subject)"),method="llrt")

```
summary(clas2.bic)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Classical ~ Instrument + Voice + (1 | Subject) + (0 + Instrument |
##     Subject) + (0 + 1 | Subject)
##    Data: data
##
## REML criterion at convergence: 10179.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3989 -0.5894  0.0101  0.5528  5.3764
##
## Random effects:
##  Groups    Name             Variance Std.Dev. Corr
##  Subject   (Intercept)      0.4380   0.6618
##  Subject.1 Instrumentguitar 1.5592   1.2487
##            Instrumentpiano  1.2972   1.1390    0.44
##            Instrumentstring 1.0838   1.0411   -0.33  0.19
##  Subject.2 (Intercept)      0.7416   0.8611
##  Residual                   2.8978   1.7023
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       4.54085    0.21199  21.420
## Instrumentpiano   1.36703    0.17313   7.896
## Instrumentstring  3.12803    0.23828  13.127
## Voicepar3rd      -0.41087    0.08355  -4.918
## Voicepar5th      -0.36793    0.08350  -4.407
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts Vcpr3r
## Instrumntpn -0.458
## Instrmntstr -0.630  0.632
## Voicepar3rd -0.196 -0.001    -0.001
## Voicepar5th -0.197 -0.001     0.000     0.500
## convergence code: 0
## Model failed to converge with max|grad| = 0.0274357 (tol = 0.002, component 1)
```

```r
summary(clas2.llrt)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject) + (0 +
##     Instrument | Subject) + (0 + Harmony | Subject) + Harmony:Voice
##    Data: music.1
##
## REML criterion at convergence: 9916.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7397 -0.5713  0.0174  0.5572  6.1551
##
## Random effects:
##  Groups    Name             Variance Std.Dev. Corr
```

```
##  Subject   (Intercept)      1.46383  1.20989
##  Subject.1 Instrumentguitar 1.25254  1.11917
##            Instrumentpiano  1.12718  1.06169   0.30
##            Instrumentstring 1.13894  1.06721  -0.49  0.11
##  Subject.2 HarmonyI-IV-V    0.00000  0.00000
##            HarmonyI-V-IV    0.06319  0.25137    NaN
##            HarmonyI-V-VI    1.44367  1.20153    NaN -0.06
##            HarmonyIV-I-V    0.00442  0.06648    NaN  0.50 -0.89
##  Residual                  2.43043  1.55898
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)               4.25308    0.22888  18.583
## Instrumentpiano           1.36898    0.17228   7.946
## Instrumentstring          3.12676    0.23816  13.129
## HarmonyI-V-IV             0.15464    0.15608   0.991
## HarmonyI-V-VI             1.13767    0.20985   5.421
## HarmonyIV-I-V            -0.13399    0.15271  -0.877
## Voicepar3rd              -0.27038    0.15256  -1.772
## Voicepar5th              -0.23557    0.15292  -1.541
## HarmonyI-V-IV:Voicepar3rd -0.36471    0.21622  -1.687
## HarmonyI-V-VI:Voicepar3rd -0.68048    0.21646  -3.144
## HarmonyIV-I-V:Voicepar3rd  0.48658    0.21612   2.251
## HarmonyI-V-IV:Voicepar5th -0.18997    0.21664  -0.877
## HarmonyI-V-VI:Voicepar5th -0.42471    0.21639  -1.963
## HarmonyIV-I-V:Voicepar5th  0.07548    0.21600   0.349

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```r
summary(clas2.t)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject) + (0 +
##     Instrument | Subject) + (0 + Harmony | Subject) + Harmony:Voice
##    Data: music.1
##
## REML criterion at convergence: 9916.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7397 -0.5713  0.0174  0.5572  6.1551
##
## Random effects:
##  Groups    Name             Variance Std.Dev. Corr
##  Subject   (Intercept)      1.46383  1.20989
##  Subject.1 Instrumentguitar 1.25254  1.11917
```

```
##            Instrumentpiano  1.12718  1.06169   0.30
##            Instrumentstring 1.13894  1.06721  -0.49  0.11
##  Subject.2 HarmonyI-IV-V    0.00000  0.00000
##            HarmonyI-V-IV    0.06319  0.25137    NaN
##            HarmonyI-V-VI    1.44367  1.20153    NaN -0.06
##            HarmonyIV-I-V    0.00442  0.06648    NaN  0.50 -0.89
##  Residual                   2.43043  1.55898
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                 4.25308    0.22888  18.583
## Instrumentpiano             1.36898    0.17228   7.946
## Instrumentstring            3.12676    0.23816  13.129
## HarmonyI-V-IV               0.15464    0.15608   0.991
## HarmonyI-V-VI               1.13767    0.20985   5.421
## HarmonyIV-I-V              -0.13399    0.15271  -0.877
## Voicepar3rd                -0.27038    0.15256  -1.772
## Voicepar5th                -0.23557    0.15292  -1.541
## HarmonyI-V-IV:Voicepar3rd -0.36471    0.21622  -1.687
## HarmonyI-V-VI:Voicepar3rd -0.68048    0.21646  -3.144
## HarmonyIV-I-V:Voicepar3rd  0.48658    0.21612   2.251
## HarmonyI-V-IV:Voicepar5th -0.18997    0.21664  -0.877
## HarmonyI-V-VI:Voicepar5th -0.42471    0.21639  -1.963
## HarmonyIV-I-V:Voicepar5th  0.07548    0.21600   0.349
```

```
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Constructed lmer.clas3, lmer.clas4, lmer.clas4 based on the fitLMER.fnc() results above. Random slope on Instrument and Harmony was agreed by two out of three approaches. AIC and BIC have decreased after adding the two random slopes. lmer.clas5 gives the lowest AIC and BIC.

```
lmer.clas3 <- lmer(Classical ~ Instrument + Voice + (1 + Instrument | Subject), data = music.1,REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00202439
## (tol = 0.002, component 1)
```

```
lmer.clas4 <- lmer(Classical ~ Instrument + Voice * Harmony +  (1 + Instrument | Subject), data = music

#best model
lmer.clas5 <-  lmer(Classical ~ Instrument + Harmony  * Voice + ( Instrument + Harmony | Subject), data
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5b <- lmer(Classical ~ Instrument + Harmony  * Voice + (1 + Instrument + Harmony + Voice | Sub
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5c <- lmer(Classical ~ Instrument + Harmony  * Voice + (1 + Instrument + Voice | Subject), data
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas2, lmer.clas3, lmer.clas4, lmer.clas5,lmer.clas5b)
```

```
## Data: music.1
## Models:
## lmer.clas3: Classical ~ Instrument + Voice + (1 + Instrument | Subject)
## lmer.clas2: Classical ~ Instrument + Harmony * Voice + (1 | Subject)
## lmer.clas4: Classical ~ Instrument + Voice * Harmony + (1 + Instrument |
## lmer.clas4:     Subject)
## lmer.clas5: Classical ~ Instrument + Harmony * Voice + (Instrument + Harmony |
## lmer.clas5:     Subject)
## lmer.clas5b: Classical ~ Instrument + Harmony * Voice + (1 + Instrument +
## lmer.clas5b:     Harmony + Voice | Subject)
##             Df     AIC   BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## lmer.clas3  12 10192.1 10262 -5084.1  10168.1
## lmer.clas2  16 10458.1 10551 -5213.1  10426.1   0.000      4      1.000
## lmer.clas4  21 10085.0 10207 -5021.5  10043.0 383.143      5     <2e-16
## lmer.clas5  36  9937.3 10147 -4932.6   9865.3 177.690     15     <2e-16
## lmer.clas5b 51  9949.5 10246 -4923.8   9847.5  17.751     15      0.276
##
## lmer.clas3
## lmer.clas2
## lmer.clas4  ***
## lmer.clas5  ***
## lmer.clas5b
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking whether for Voice, Contrary has the strongest association with Classical. Found that contrary does have larger coefficient than par5th and par3rd.

```r
#failed to converge
lmer.clas5.contrary <- lmer(Classical ~ -1 + Voice * Harmony + Instrument + (1 + Instrument + Harmony |
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0301093
## (tol = 0.002, component 1)
```

```r
lm.clas5.contrary.b <- lm(Classical ~ -1 + Voice * Harmony + Instrument , data = music.1)
summary(lmer.clas5.contrary)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
```

```
## Classical ~ -1 + Voice * Harmony + Instrument + (1 + Instrument +
##     Harmony | Subject)
##    Data: music.1
##
##      AIC      BIC   logLik deviance df.resid
##   9937.0  10146.5  -4932.5   9865.0     2457
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7175 -0.5819  0.0197  0.5655  6.1294
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      2.543477 1.59483
##           Instrumentpiano  1.636692 1.27933  -0.39
##           Instrumentstring 3.511282 1.87384  -0.57  0.66
##           HarmonyI-V-IV    0.041655 0.20410   0.70 -0.67 -0.44
##           HarmonyI-V-VI    1.587792 1.26008  -0.05 -0.27 -0.42  0.21
##           HarmonyIV-I-V    0.008963 0.09467   0.19 -0.38  0.14  0.27  0.13
##  Residual                  2.417678 1.55489
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                            Estimate Std. Error t value
## Voicecontrary               4.25225    0.22326  19.047
## Voicepar3rd                 3.98184    0.22337  17.826
## Voicepar5th                 4.01580    0.22347  17.970
## HarmonyI-V-IV               0.15517    0.15468   1.003
## HarmonyI-V-VI               1.13872    0.21441   5.311
## HarmonyIV-I-V              -0.13338    0.15252  -0.874
## Instrumentpiano             1.37007    0.17120   8.003
## Instrumentstring            3.12747    0.23656  13.221
## Voicepar3rd:HarmonyI-V-IV -0.36488    0.21565  -1.692
## Voicepar5th:HarmonyI-V-IV -0.18920    0.21607  -0.876
## Voicepar3rd:HarmonyI-V-VI -0.68009    0.21589  -3.150
## Voicepar5th:HarmonyI-V-VI -0.42576    0.21582  -1.973
## Voicepar3rd:HarmonyIV-I-V  0.48537    0.21555   2.252
## Voicepar5th:HarmonyIV-I-V  0.07525    0.21544   0.349

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## Model failed to converge with max|grad| = 0.0301093 (tol = 0.002, component 1)
```

```
summary(lm.clas5.contrary.b)
```

```
##
## Call:
## lm(formula = Classical ~ -1 + Voice * Harmony + Instrument, data = music.1)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -6.8570 -1.7245  0.0228  1.6715 11.7530 
## 
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)    
## Voicecontrary               4.24840    0.17127  24.806  < 2e-16 ***
## Voicepar3rd                 3.97389    0.17144  23.179  < 2e-16 ***
## Voicepar5th                 4.01577    0.17186  23.367  < 2e-16 ***
## HarmonyI-V-IV               0.15288    0.22506   0.679   0.4970    
## HarmonyI-V-VI               1.14044    0.22478   5.074 4.19e-07 ***
## HarmonyIV-I-V              -0.13397    0.22424  -0.597   0.5503    
## Instrumentpiano             1.37398    0.11277  12.184  < 2e-16 ***
## Instrumentstring            3.13256    0.11209  27.947  < 2e-16 ***
## Voicepar3rd:HarmonyI-V-IV -0.35922    0.31770  -1.131   0.2583    
## Voicepar5th:HarmonyI-V-IV -0.19140    0.31846  -0.601   0.5479    
## Voicepar3rd:HarmonyI-V-VI -0.68392    0.31808  -2.150   0.0316 *  
## Voicepar5th:HarmonyI-V-VI -0.43173    0.31789  -1.358   0.1746    
## Voicepar3rd:HarmonyIV-I-V  0.48855    0.31770   1.538   0.1242    
## Voicepar5th:HarmonyIV-I-V  0.06817    0.31751   0.215   0.8300    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.292 on 2479 degrees of freedom
## Multiple R-squared:  0.871,  Adjusted R-squared:  0.8703 
## F-statistic:  1196 on 14 and 2479 DF,  p-value: < 2.2e-16
```

Checking for Harmony, whether it is associated with if people is familiar with one or the other (or both) of the Pachelbel rants/comedy bits. Result suggest that interaction between harmony and KnowRob/KnowAxis should not be added to model.

```
lmer.clas5.sub <-  lmer(Classical ~ Instrument + Harmony  * Voice + (1 + Instrument + Harmony | Subject
```

```
## boundary (singular) fit: see ?isSingular
```

```
lmer.clas5.harmR <-  lmer(Classical ~ Instrument + Harmony  * Voice + KnowRob +KnowRob: Harmony  + (1 +
```

```
## boundary (singular) fit: see ?isSingular
```

```
lmer.clas5.harmA <-  lmer(Classical ~ Instrument + Harmony  * Voice + KnowAxis +KnowAxis: Harmony  + (1
```

```
## boundary (singular) fit: see ?isSingular
```

```
lmer.clas5.harmB <-  lmer(Classical ~ Instrument + Harmony  * Voice + KnowRob + KnowAxis +KnowAxis: Har
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(lmer.clas5.sub, lmer.clas5.harmR ,lmer.clas5.harmA, lmer.clas5.harmB )
```

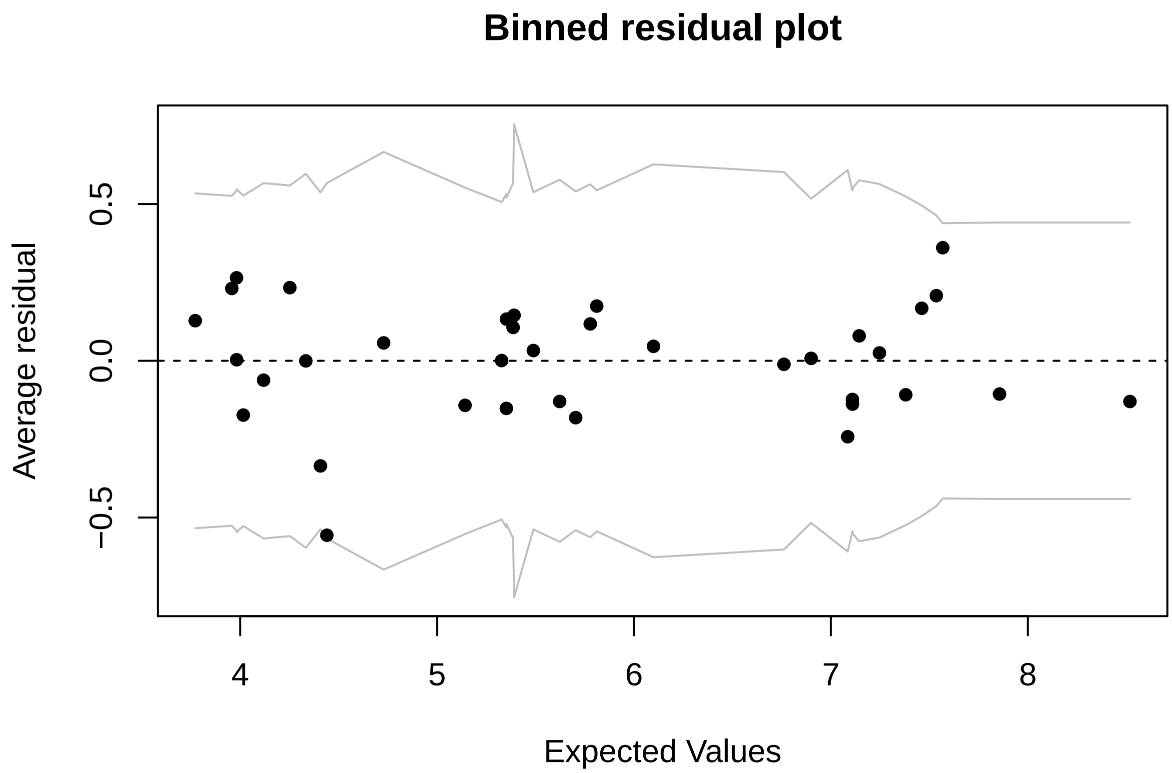
```
## Data: music.sub
## Models:
## lmer.clas5.sub: Classical ~ Instrument + Harmony * Voice + (1 + Instrument +
## lmer.clas5.sub:     Harmony | Subject)
## lmer.clas5.harmR: Classical ~ Instrument + Harmony * Voice + KnowRob + KnowRob:Harmony +
## lmer.clas5.harmR:     (1 + Instrument + Harmony | Subject)
## lmer.clas5.harmA: Classical ~ Instrument + Harmony * Voice + KnowAxis + KnowAxis:Harmony +
## lmer.clas5.harmA:     (1 + Instrument + Harmony | Subject)
## lmer.clas5.harmB: Classical ~ Instrument + Harmony * Voice + KnowRob + KnowAxis +
## lmer.clas5.harmB:     KnowAxis:Harmony + KnowAxis:Harmony + (1 + Instrument + Harmony |
## lmer.clas5.harmB:     Subject)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas5.sub   36 6190.1 6382.3 -3059.0   6118.1
## lmer.clas5.harmR 44 6196.1 6431.1 -3054.1   6108.1 9.9460      8
## lmer.clas5.harmA 44 6197.6 6432.5 -3054.8   6109.6 0.0000      0
## lmer.clas5.harmB 46 6200.3 6446.0 -3054.2   6108.3 1.2467      2
##                  Pr(>Chisq)
## lmer.clas5.sub
## lmer.clas5.harmR     0.2688
## lmer.clas5.harmA     1.0000
## lmer.clas5.harmB     0.5361
```

Residual analysis for best fit (lmer.clas5):

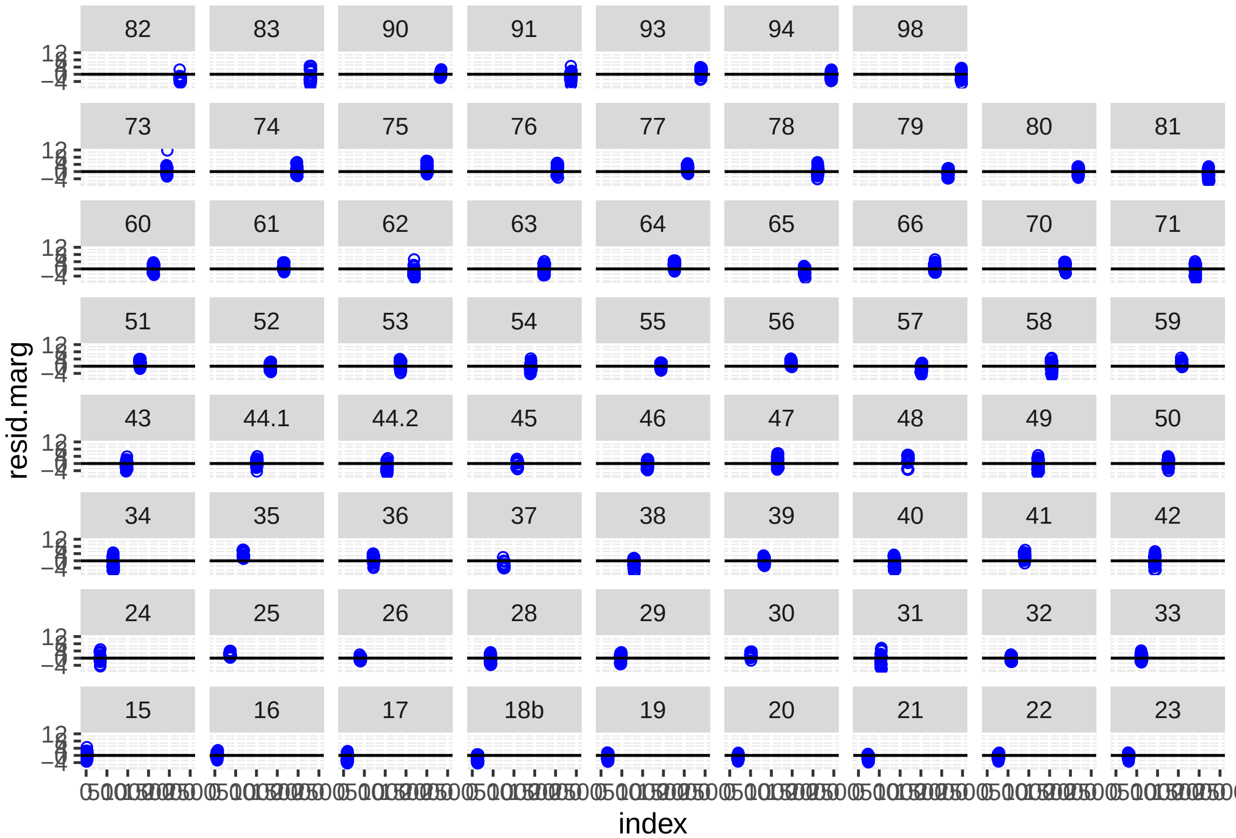
```r
resid.marg <- r.marg(lmer.clas5)
resid.cond <- r.cond(lmer.clas5)
resid.reff <- r.reff(lmer.clas5)

fit.marg <- yhat.marg(lmer.clas5)
fit.cond <- yhat.cond(lmer.clas5)
fit.reff <- yhat.reff(lmer.clas5)
```

. Below are the maginal residual plots. In binned plot we see all points falls in between the +-2 SE bound which is preferred. In the second plot we see the residuals for each subject are centered at 0, which is good. Finally Q-Q plot shows slightly short tails on both ends

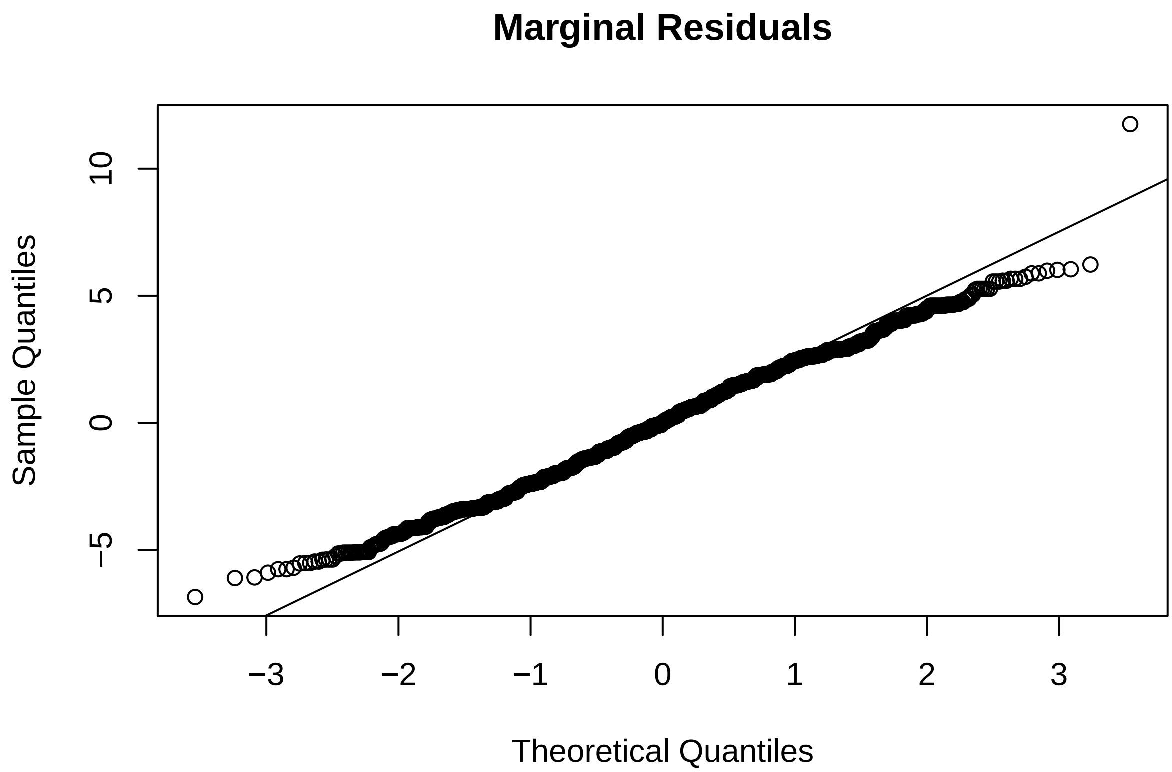
```r
binnedplot(fit.marg,resid.marg)
```




```
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.marg,music.1$Subject)
names(new.data) <- c("index","resid.marg","Subject")
ggplot(new.data,aes(x=index,y=resid.marg)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```



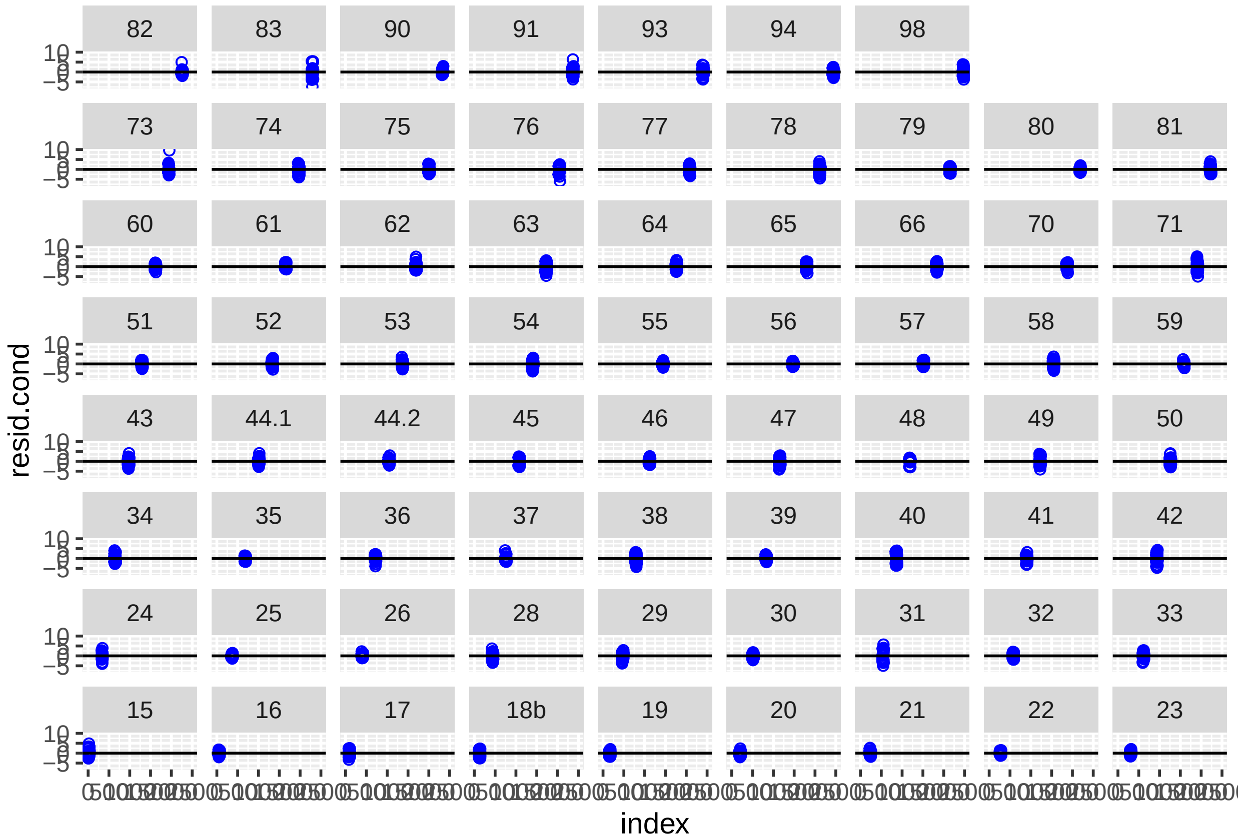

```r
qqnorm(resid.marg,main="Marginal Residuals")
qqline(resid.marg)
```



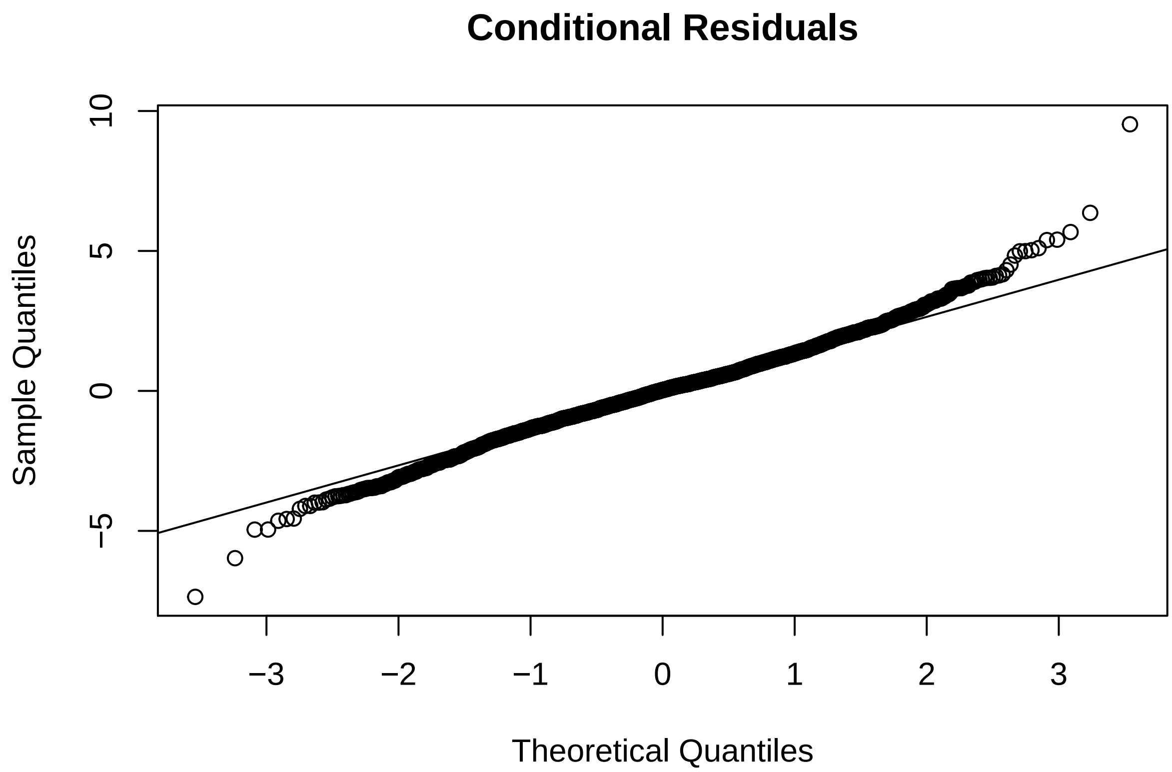
Below are the conditional residual plots. In binned plot we see above 95% of points falls in between the +-2 SE bound and there are no patterns displayed by the residuals, which is preferred. In the second plot we see the residuals for each subject are centered at 0, which is good. Finally Q-Q plot shows slightly short tails on both ends but the issue is not bad. In general this indicates our model is a good fit.

```
binnedplot(fit.cond,resid.cond)
```


```
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.cond,music.1$Subject)
names(new.data) <- c("index","resid.cond","Subject")
ggplot(new.data,aes(x=index,y=resid.cond)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```

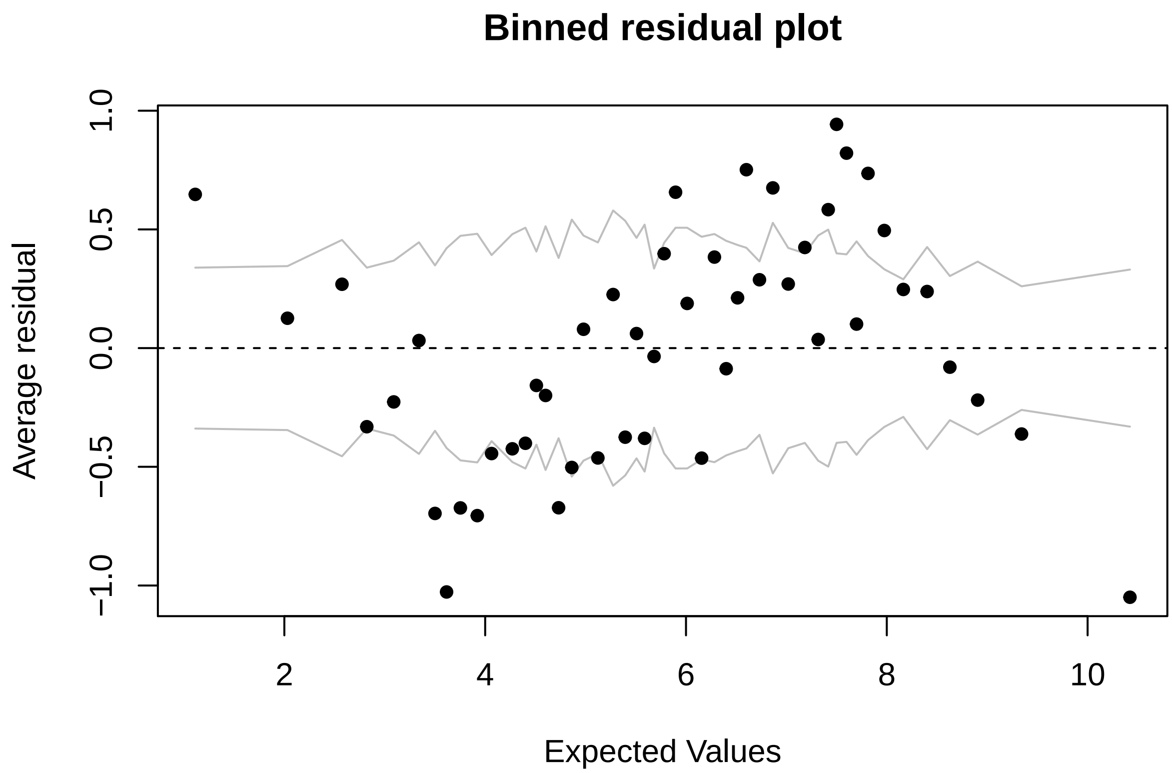



```r
qqnorm(resid.cond,main="Conditional Residuals")
qqline(resid.cond)
```

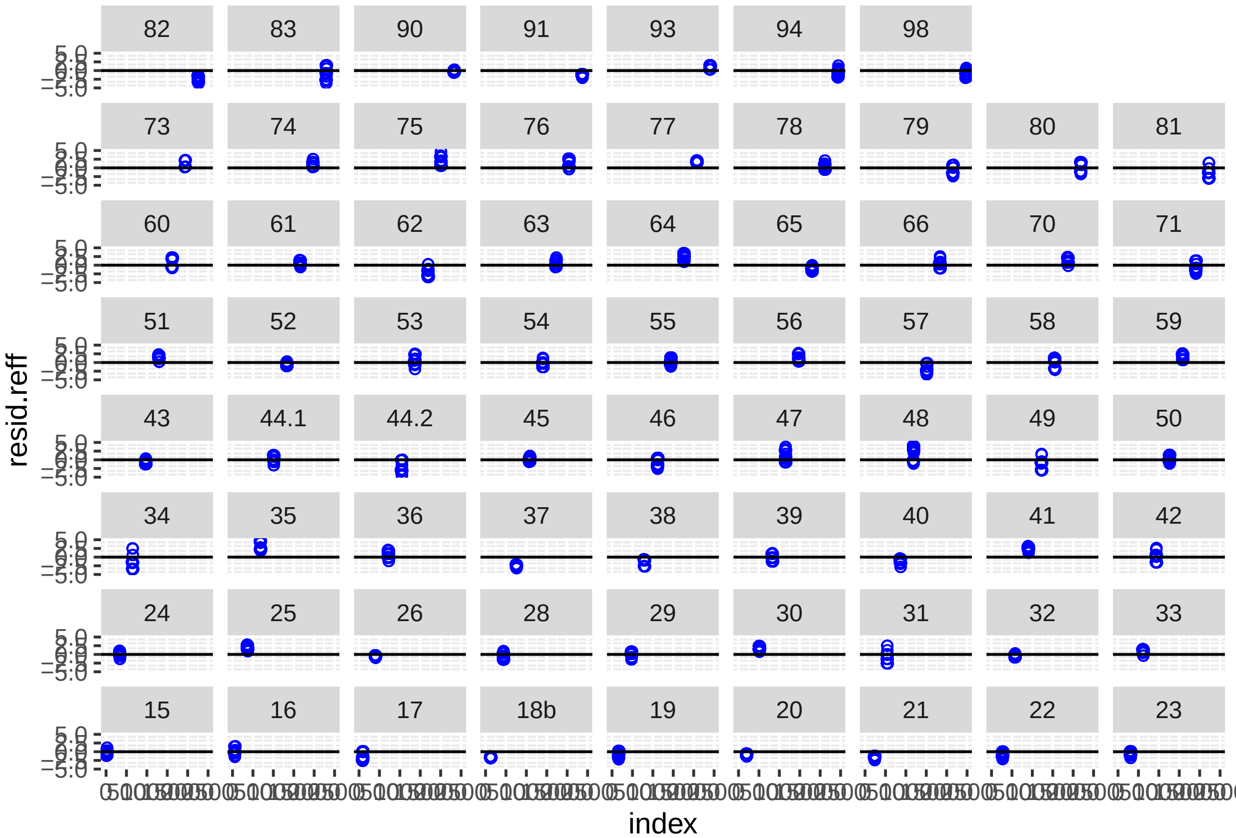


Finally for the random effect residuals, which we should be mainly concentrating on Q-Q plot. In Q-Q plot we again see slightly short tails on both ends.

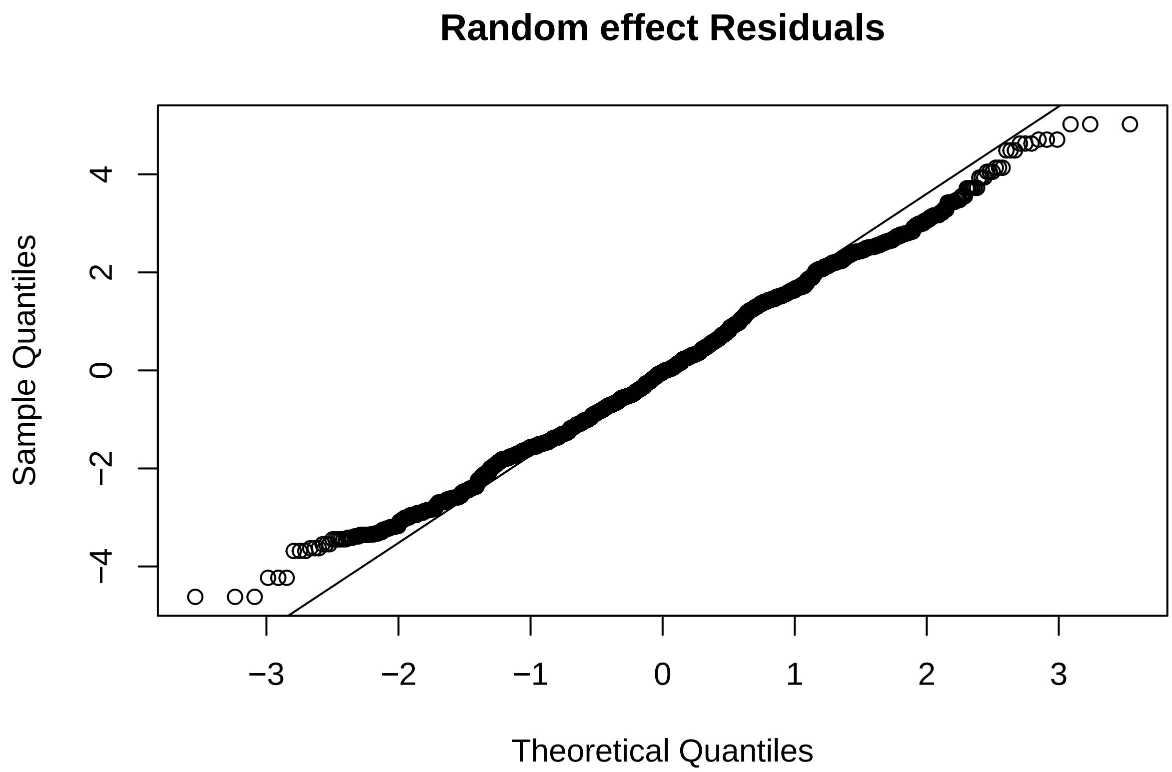
```
binnedplot(fit.reff,resid.reff)
```




```
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.reff,music.1$Subject)
names(new.data) <- c("index","resid.reff","Subject")
ggplot(new.data,aes(x=index,y=resid.reff)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```




```r
qqnorm(resid.reff,main="Random effect Residuals")
qqline(resid.reff)
```



Fitting Popular with Instrument, Harmony, Voice

```
lm.pop.base <- lm(Popular ~ Instrument + Harmony + Voice, data = music.1)
lm.pop1 <- lm(Popular ~ Instrument*Harmony*Voice, data = music.1)
lm.pop1.aic <- stepAIC(lm.pop1,trace=0, direction="both")
summary(lm.pop1.aic)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony, data = music.1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6930 -1.6930  0.1867  1.4927 13.2824
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.69301    0.11055  60.541   <2e-16 ***
## Instrumentpiano  -0.95186    0.11104  -8.572   <2e-16 ***
## Instrumentstring -2.61166    0.11037 -23.662   <2e-16 ***
## HarmonyI-V-IV    -0.02351    0.12784  -0.184   0.8541
## HarmonyI-V-VI    -0.26805    0.12784  -2.097   0.0361 *
## HarmonyIV-I-V    -0.18575    0.12774  -1.454   0.1460
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.257 on 2487 degrees of freedom
## Multiple R-squared:  0.1891, Adjusted R-squared:  0.1875
## F-statistic:   116 on 5 and 2487 DF,  p-value: < 2.2e-16
```

lm.pop.a is fitted based on the output from stepAIC, which is lm.pop1.aic. We have added Harmony back in even though stepAIC have decided that it should be dropped

```
lm.pop.a <- lm(Popular ~ Instrument + Harmony + Voice, data = music.1)
anova(lm.pop.base, lm.pop.a,lm.pop1)
```

```
## Analysis of Variance Table
##
## Model 1: Popular ~ Instrument + Harmony + Voice
## Model 2: Popular ~ Instrument + Harmony + Voice
## Model 3: Popular ~ Instrument * Harmony * Voice
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1   2485 12656
## 2   2485 12656  0      0.00
## 3   2457 12517 28    139.58 0.9785 0.4971
```

Testing for random intercept: both AIC and BIC dropped significantly after adding random intercept

```
lmer.pop1 <- lmer(Popular ~ Instrument + Harmony + Voice + (1|Subject), data = music.1,REML=F)
summary(lmer.pop1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: music.1
##
##      AIC      BIC   logLik deviance df.resid
##  10430.3  10488.5  -5205.1  10410.3     2483
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7302 -0.6447  0.0394  0.6630  5.6235
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.542    1.242
##  Residual             3.522    1.877
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        6.57768    0.18247  36.048
## Instrumentpiano   -0.94546    0.09241 -10.231
## Instrumentstring  -2.60670    0.09181 -28.393
## HarmonyI-V-IV     -0.02495    0.10630  -0.235
## HarmonyI-V-VI     -0.27227    0.10630  -2.561
```

```
## HarmonyIV-I-V    -0.18616    0.10621  -1.753
## Voicepar3rd       0.17050    0.09211   1.851
## Voicepar5th       0.16517    0.09205   1.794
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.251
## Instrmntstr -0.252  0.498
## HrmnyI-V-IV -0.290  0.001    -0.001
## HrmnyI-V-VI -0.290  0.001    -0.001     0.499
## HrmnyIV-I-V -0.291 -0.001    -0.001     0.499  0.499
## Voicepar3rd -0.252 -0.001    -0.001    -0.002  0.001  0.002
## Voicepar5th -0.251 -0.001     0.000    -0.002 -0.003 -0.001  0.500
```

```
anova(lmer.pop1,lm.pop.a)
```

```
## Data: music.1
## Models:
## lm.pop.a: Popular ~ Instrument + Harmony + Voice
## lmer.pop1: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
##           Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm.pop.a   9 11143 11196 -5562.6    11125
## lmer.pop1 10 10430 10488 -5205.1    10410 714.85      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for random slopes:All three approaches suggests adding random slope on Instrument

lmer.pop1.bic <- fitLMER.fnc(lmer.pop1,method = “BIC”,ran.effects = c(“(Instrument|Subject)”,“(Harmony|Subject)”,“(Voi
lmer.pop1.t <- fitLMER.fnc(lmer.pop1,method = “t”,ran.effects = c(“(Instrument|Subject)”,“(Harmony|Subject)”,“(Voice|Su
lmer.pop1.llrt <- fitLMER.fnc(lmer.pop1,method = “AIC”,ran.effects = c(“(Instrument|Subject)”,“(Harmony|Subject)”,“(Voi

```
anova(lmer.pop1.bic,lmer.pop1.t,lmer.pop1.llrt)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
## Models:
## lmer.pop1.bic: Popular ~ Instrument + (1 | Subject) + (Instrument | Subject)
## lmer.pop1.t: Popular ~ Instrument + (1 | Subject) + (Instrument | Subject)
## lmer.pop1.llrt: Popular ~ Instrument + (1 | Subject) + (Instrument | Subject)
##                Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.pop1.bic  11 10106 10170  -5042    10084
## lmer.pop1.t    11 10106 10170  -5042    10084     0      0          1
## lmer.pop1.llrt 11 10106 10170  -5042    10084     0      0          1
```

```
lmer.pop2 <- lmer(Popular ~ Instrument + Harmony + Voice + (1 + Instrument| Subject),data = music.1, REM
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00345307
## (tol = 0.002, component 1)
```

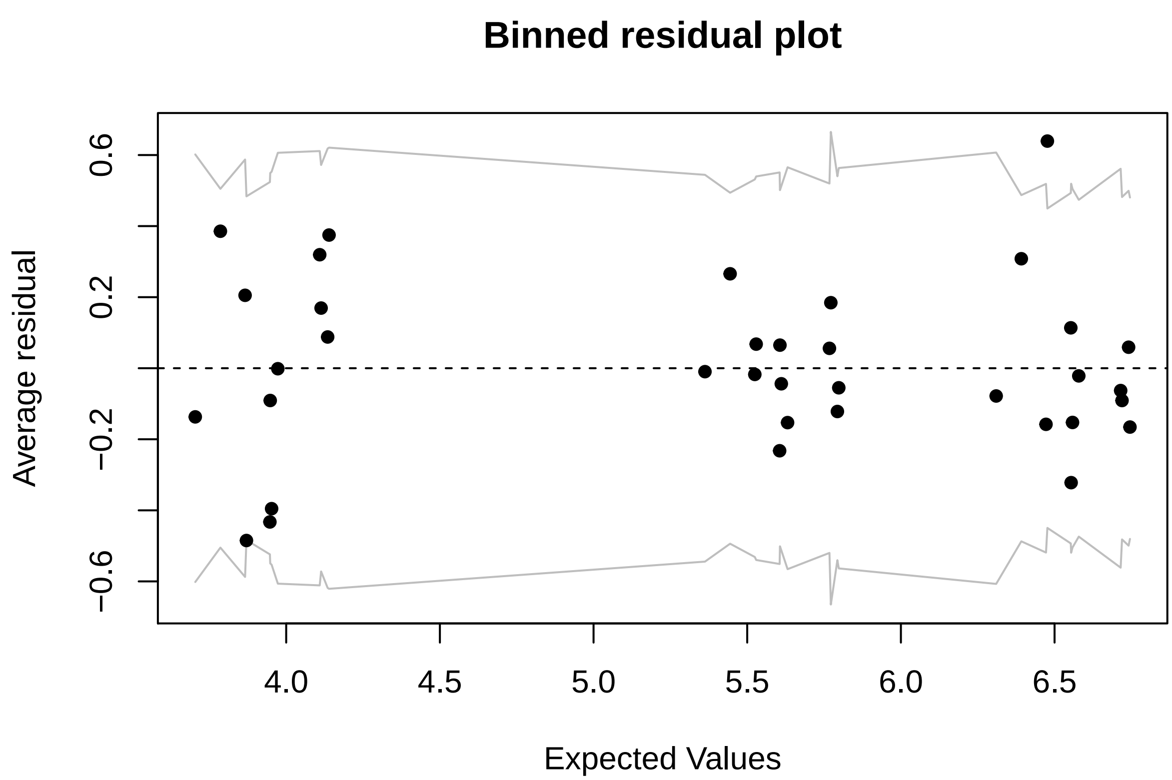
```r
anova(lmer.pop1,lmer.pop2)
```

```
## Data: music.1
## Models:
## lmer.pop1: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.pop2: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
##           Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.pop1 10 10430 10488 -5205.1    10410
## lmer.pop2 15 10098 10185 -5033.9    10068 342.56      5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual analysis for lmer.pop2

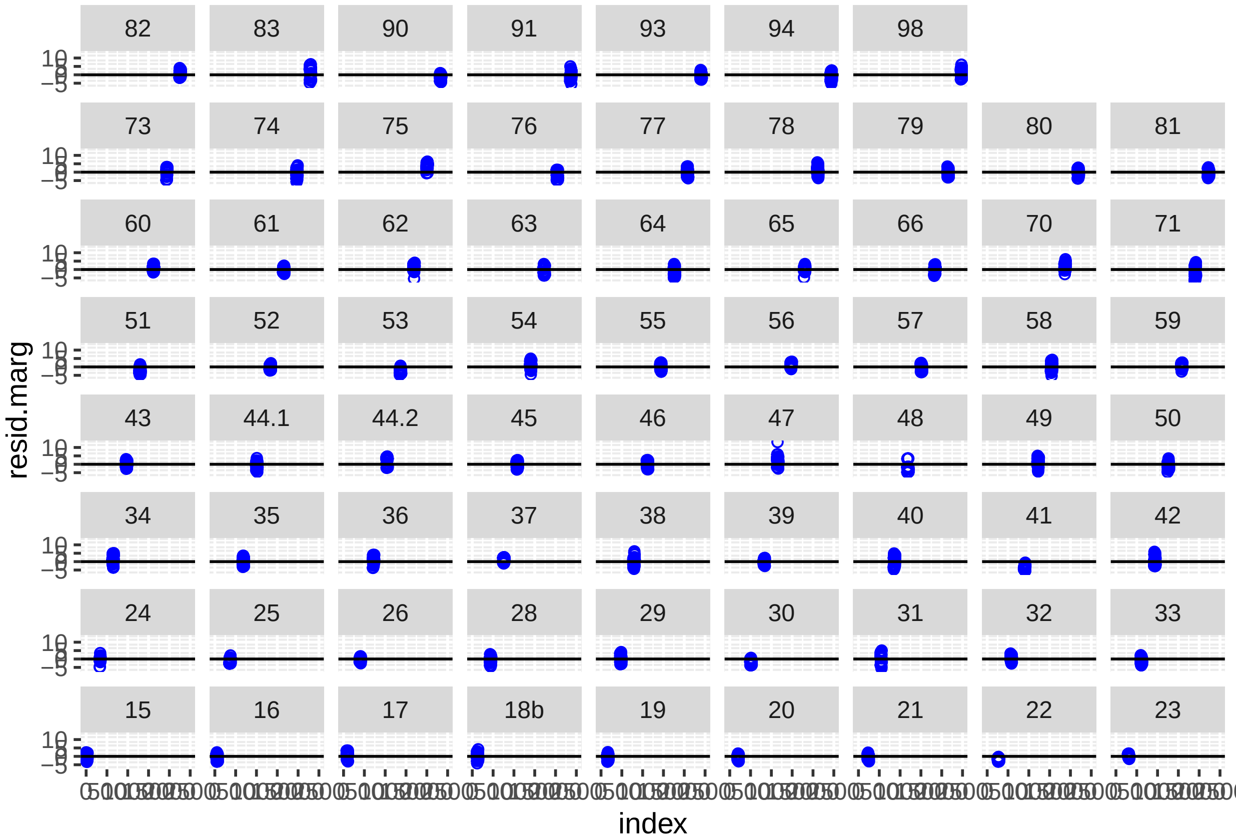
Below are the maginal residual plots. In binned plot we see all points falls in between the +-2 SE bound which is preferred. In the second plot we see the residuals for each subject are centered at 0, which is good. Finally Q-Q plot shows slightly short tails on both ends.

```r
binnedplot(fit.marg,resid.marg)
```

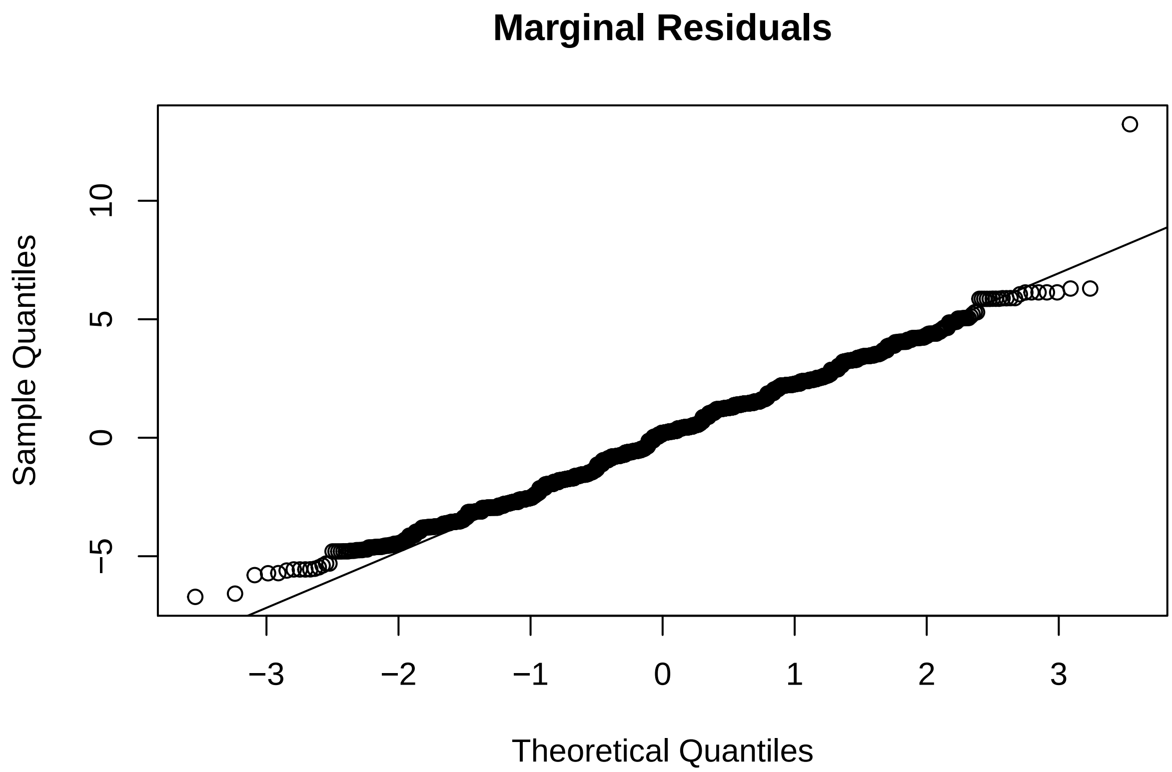



```r
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.marg,music.1$Subject)
names(new.data) <- c("index","resid.marg","Subject")
```

```r
ggplot(new.data,aes(x=index,y=resid.marg)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```

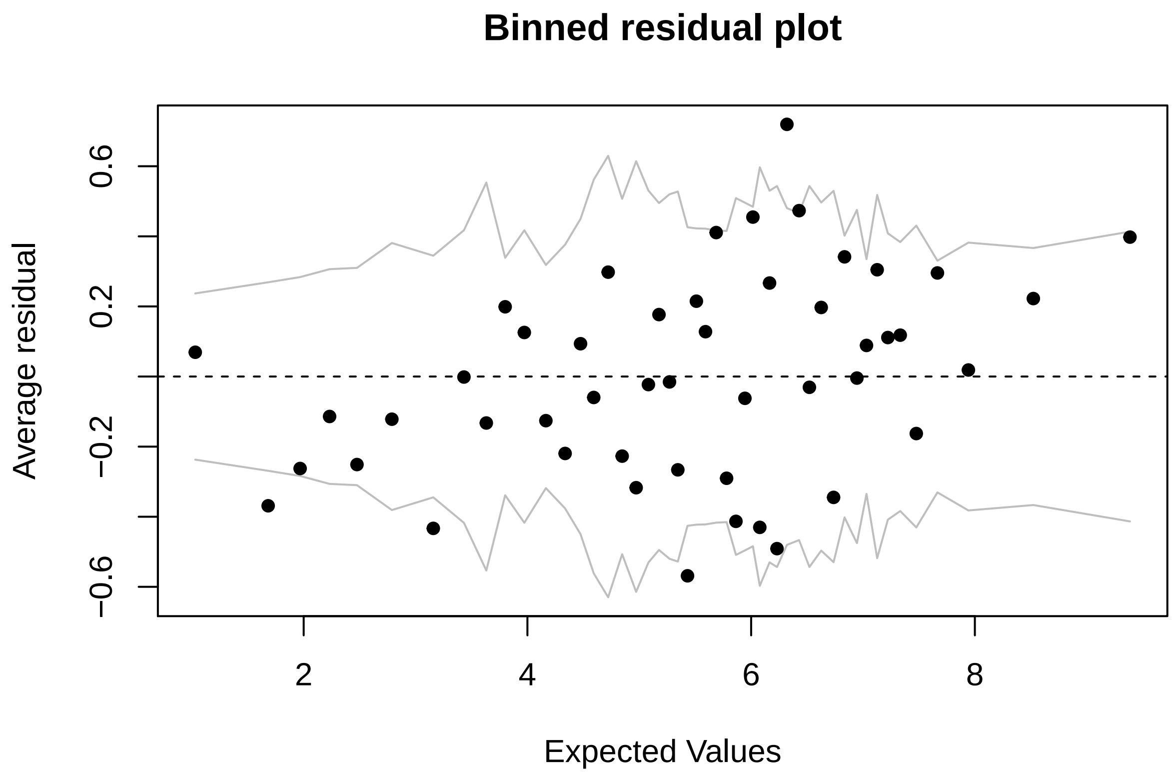



```r
qqnorm(resid.marg,main="Marginal Residuals")
qqline(resid.marg)
```

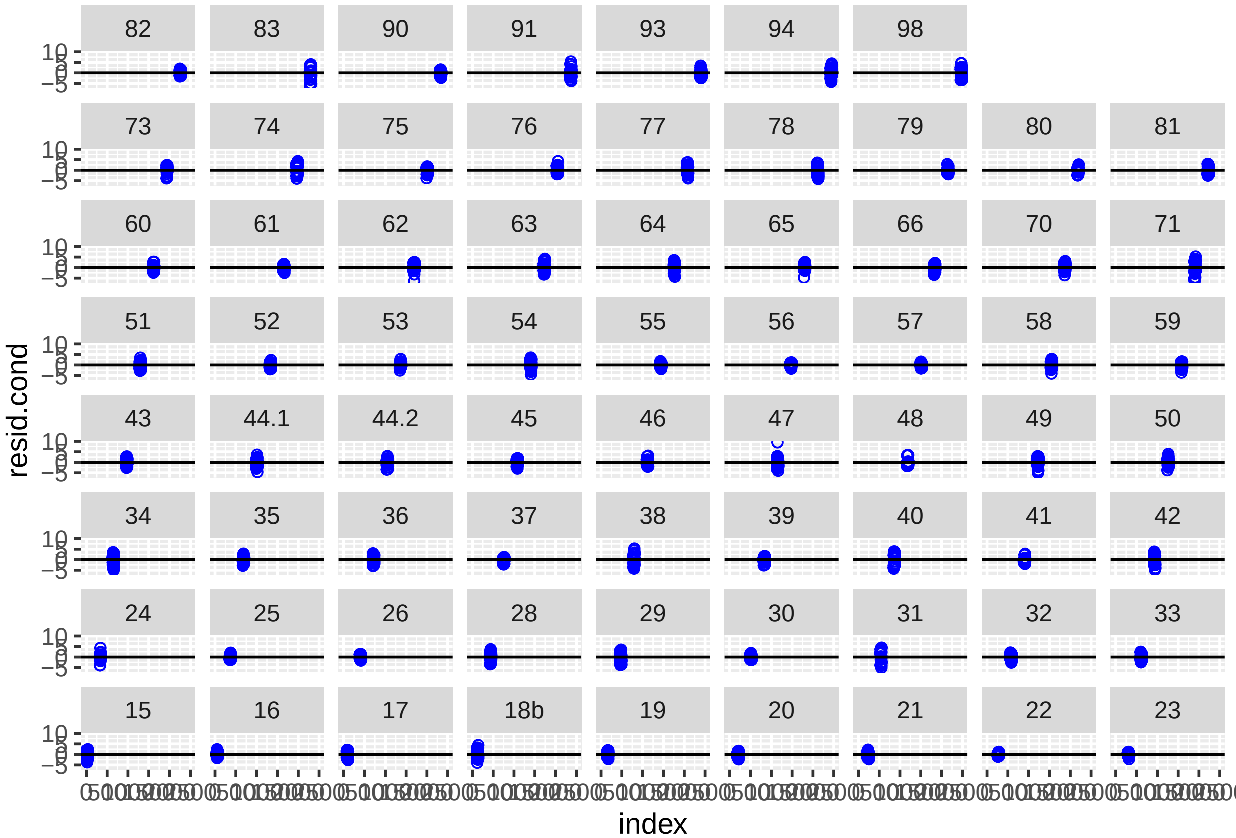


Below are the conditional residual plots. In binned plot we see above 95% of points falls in between the +-2 SE bound and there are no patterns displayed by the residuals, which is preferred. In the second plot we see the residuals for each subject are centered at 0, which is good. Finally Q-Q plot shows slightly long tails on both ends but the issue is not bad. In general this indicates our model is a good fit.

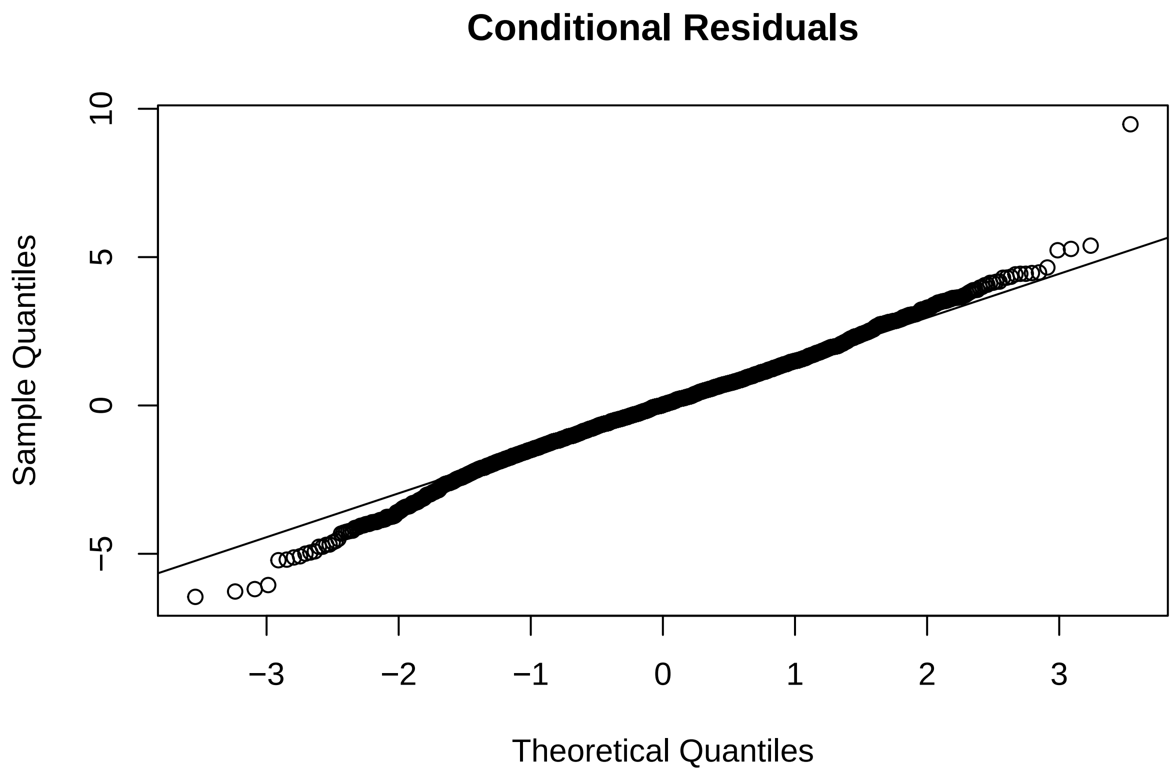
```
binnedplot(fit.cond,resid.cond)
```




```
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.cond,music.1$Subject)
names(new.data) <- c("index","resid.cond","Subject")
ggplot(new.data,aes(x=index,y=resid.cond)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```

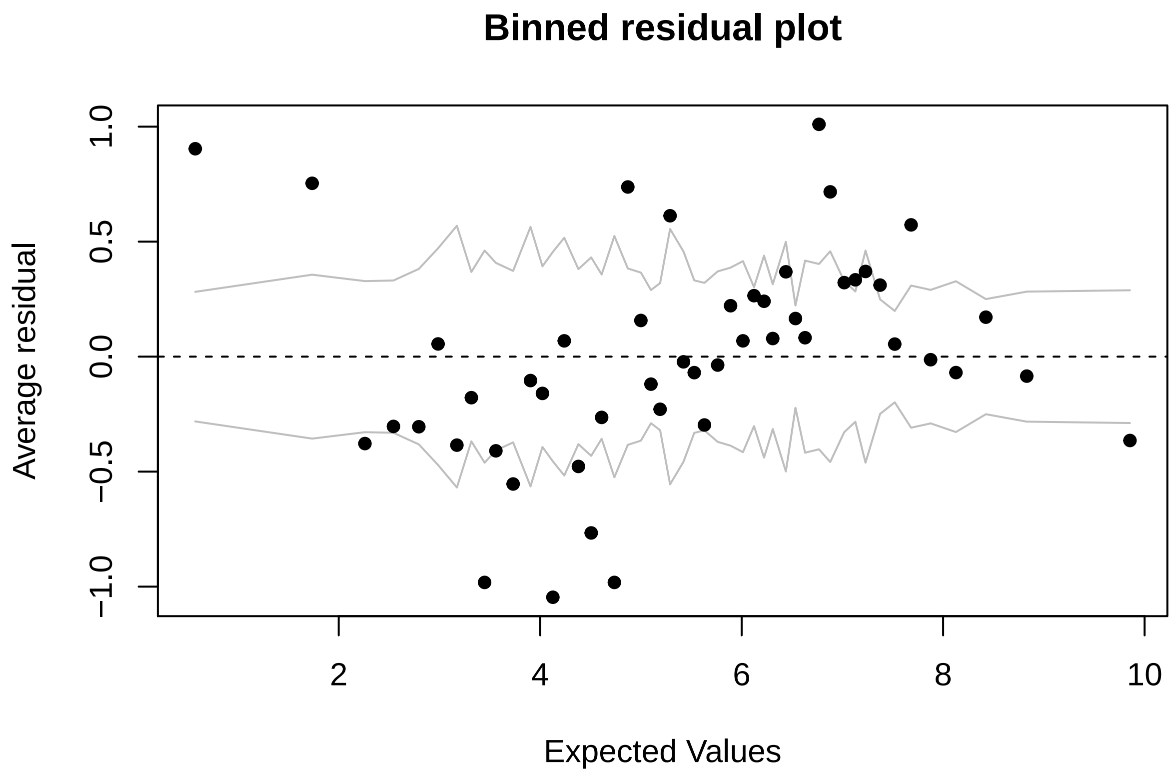



```r
qqnorm(resid.cond,main="Conditional Residuals")
qqline(resid.cond)
```

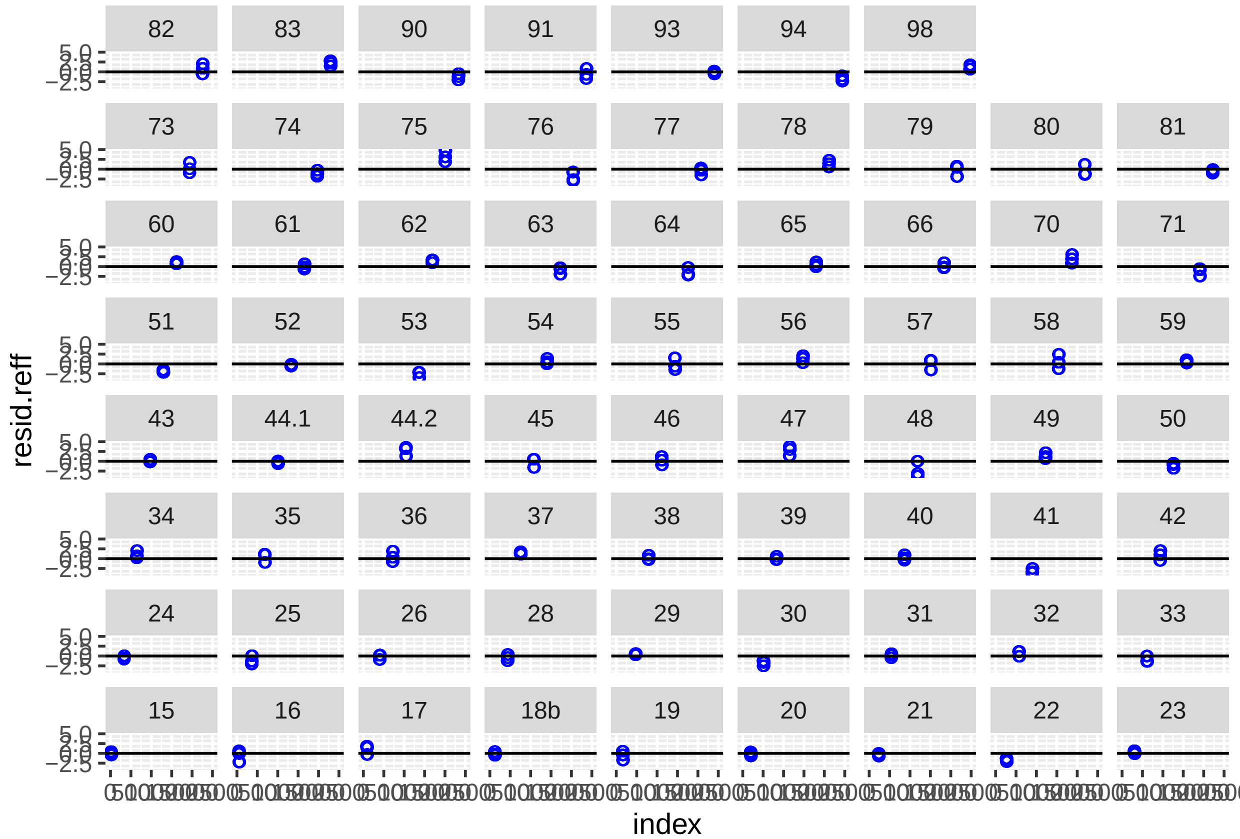


Finally for the random effect residuals, which we should be mainly concentrating on Q-Q plot. In Q-Q plot we again see slightly short tails on both ends.

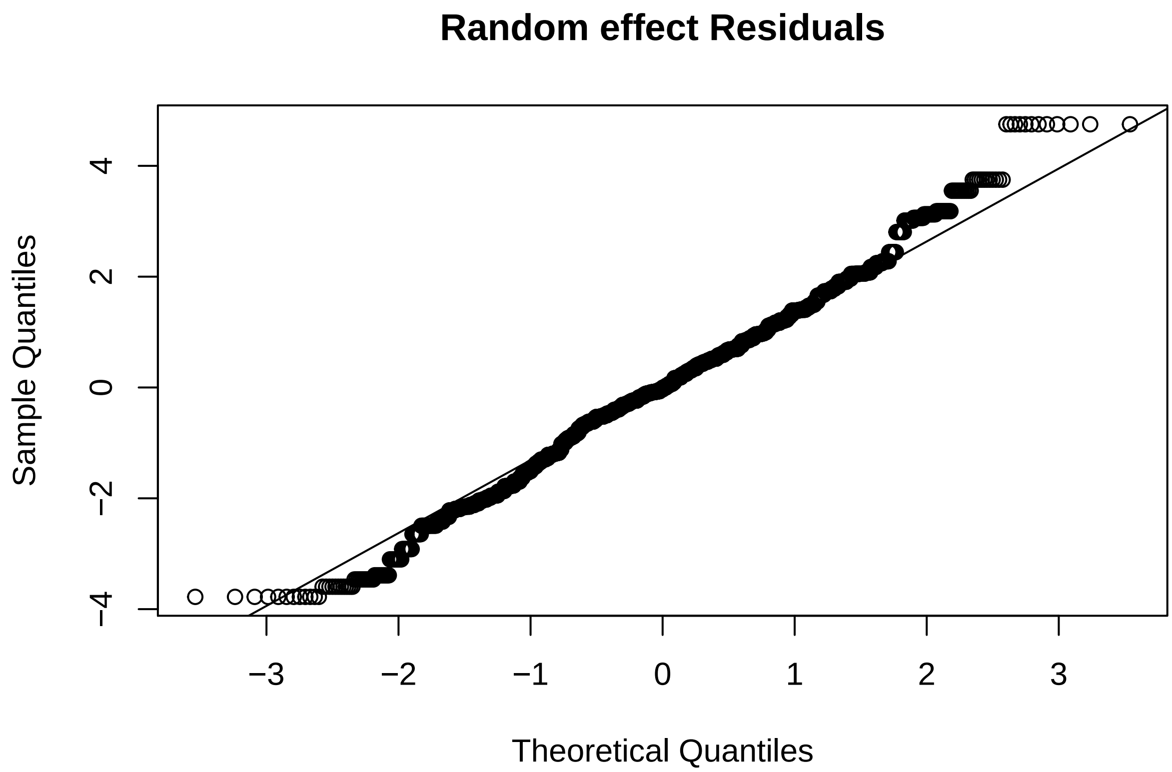
```
binnedplot(fit.reff,resid.reff)
```




```
index <- 1:dim(music.1)[1]

new.data <- data.frame(index,resid.reff,music.1$Subject)
names(new.data) <- c("index","resid.reff","Subject")
ggplot(new.data,aes(x=index,y=resid.reff)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```




```r
qqnorm(resid.reff,main="Random effect Residuals")
qqline(resid.reff)
```



Testing covariates for Classical

First use stepAIC() to select candidate fixed effects

```
lm.fixed_only <- lm(Classical ~ Instrument + Harmony * Voice  +  Selfdeclare + log_OMSI  +Instr.minus.No
lm.fixedonly.trim <- stepAIC(lm.fixed_only,direction="both",trace=0)
summary(lm.fixedonly.trim)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony + Voice + Selfdeclare +
##     log_OMSI + Instr.minus.Notes + PachListen + ClsListen + KnowAxis +
##     NoClass + PianoPlay + APTheory + Harmony:Voice, data = music.sub)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4028 -1.3603  0.0756  1.3959  5.9785
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.42483    0.53408   4.540 6.07e-06 ***
## Instrumentpiano             1.65443    0.12847  12.878  < 2e-16 ***
## Instrumentstring            3.58431    0.12806  27.989  < 2e-16 ***
## HarmonyI-V-IV               0.22290    0.25650   0.869 0.384983
```

```
## HarmonyI-V-VI             1.26973    0.25701   4.940 8.67e-07 ***
## HarmonyIV-I-V            -0.30233    0.25599  -1.181 0.237794
## Voicepar3rd              -0.31008    0.25599  -1.211 0.225982
## Voicepar5th              -0.19116    0.25650  -0.745 0.456222
## Selfdeclare2             -0.73806    0.26959  -2.738 0.006261 **
## Selfdeclare3             -0.46607    0.26035  -1.790 0.073634 .
## Selfdeclare4             -2.51939    0.48565  -5.188 2.42e-07 ***
## Selfdeclare5             -4.45644    0.57534  -7.746 1.73e-14 ***
## Selfdeclare6             -2.67589    0.97681  -2.739 0.006228 **
## log_OMSI                  0.28585    0.09552   2.993 0.002810 **
## Instr.minus.Notes         0.33189    0.05095   6.514 9.96e-11 ***
## PachListen3               0.38630    0.36682   1.053 0.292468
## PachListen4              -0.85369    0.67553  -1.264 0.206524
## PachListen5               0.44965    0.32601   1.379 0.168023
## ClsListen1               -1.04667    0.23407  -4.472 8.34e-06 ***
## ClsListen3                0.21464    0.22638   0.948 0.343212
## ClsListen4                2.30289    0.61923   3.719 0.000207 ***
## ClsListen5                0.16694    0.25598   0.652 0.514407
## KnowAxis1                 2.75370    0.48916   5.629 2.15e-08 ***
## KnowAxis5                 0.17104    0.16438   1.040 0.298277
## NoClass                  -0.33827    0.09682  -3.494 0.000490 ***
## PianoPlay1                1.30881    0.21449   6.102 1.33e-09 ***
## PianoPlay4                0.94611    0.25485   3.712 0.000213 ***
## PianoPlay5                2.99585    0.32538   9.207  < 2e-16 ***
## APTheory1                 1.10827    0.20965   5.286 1.43e-07 ***
## HarmonyI-V-IV:Voicepar3rd -0.43995    0.36239  -1.214 0.224922
## HarmonyI-V-VI:Voicepar3rd -0.72113    0.36310  -1.986 0.047210 *
## HarmonyIV-I-V:Voicepar3rd  0.75718    0.36239   2.089 0.036838 *
## HarmonyI-V-IV:Voicepar5th -0.22290    0.36309  -0.614 0.539381
## HarmonyI-V-VI:Voicepar5th -0.53981    0.36312  -1.487 0.137329
## HarmonyIV-I-V:Voicepar5th  0.32295    0.36239   0.891 0.372985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.056 on 1506 degrees of freedom
## Multiple R-squared:  0.4461, Adjusted R-squared:  0.4336
## F-statistic: 35.67 on 34 and 1506 DF,  p-value: < 2.2e-16
```

Adding the candidate fixed effects in lmer and use fitLMER.fnc() to see which ones were kept. fitLMER suggests that PianoPlay and ClsListen should be further investigates as two out of three approaches suggested these two

```
lmer.ran <- lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare +
    log_OMSI + ClsListen + KnowAxis + NoClass + PianoPlay +  Harmony:Voice + (Instrument  | Subject)+ (
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0190839
## (tol = 0.002, component 1)
```

lmer.ran.aic <- fitLMER.fnc(lmer.ran,method = "AIC") lmer.ran.t <- fitLMER.fnc(lmer.ran,method = "t") lmer.ran.llrt <- fitLMER.fnc(lmer.ran,method = "llrt")

```r
anova(lmer.ran.aic,lmer.ran.t,lmer.ran.llrt)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: music.sub
## Models:
## lmer.ran.aic: Classical ~ Instrument + Harmony + Voice + PianoPlay + (Instrument |
## lmer.ran.aic:     Subject) + (Harmony | Subject) + Harmony:Voice
## lmer.ran.t: Classical ~ Instrument + Harmony + Voice + log_OMSI + ClsListen +
## lmer.ran.t:     PianoPlay + (Instrument | Subject) + (Harmony | Subject) +
## lmer.ran.t:     Harmony:Voice
## lmer.ran.llrt: Classical ~ Instrument + Harmony + Voice + Selfdeclare + ClsListen +
## lmer.ran.llrt:     KnowAxis + PianoPlay + (Instrument | Subject) + (Harmony |
## lmer.ran.llrt:     Subject) + Harmony:Voice
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.ran.aic  34 6188.5 6370.1 -3060.2   6120.5
## lmer.ran.t    39 6191.6 6399.9 -3056.8   6113.6  6.856      5  0.2315699
## lmer.ran.llrt 45 6181.1 6421.4 -3045.6   6091.1 22.504      6  0.0009807
##
## lmer.ran.aic
## lmer.ran.t
## lmer.ran.llrt ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fitting new models based on fixed effects agreed by two approaches above (Selfdeclare, PachListen, ClsListen, PianoPlay). Based on below output, we only need to add PianoPlay.

```r
lmer.clas5sub <-  lmer(Classical ~ Instrument + Harmony + Voice +  Harmony:Voice + (Instrument  + Harmo
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5.a <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice + PianoPlay + (Instrum
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5.b <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice + PianoPlay+ Selfdecla
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5.c <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice  + Selfdeclare + PachLi
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas5.d <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice  + PianoPlay +  Selfde
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas5sub, lmer.clas5.a, lmer.clas5.b, lmer.clas5.c, lmer.clas5.d )
```

```
## Data: music.sub
## Models:
## lmer.clas5sub: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (Instrument +
## lmer.clas5sub:     Harmony | Subject)
## lmer.clas5.a: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas5.a:     (Instrument + Harmony | Subject)
## lmer.clas5.b: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas5.b:     Selfdeclare + (Instrument + Harmony | Subject)
## lmer.clas5.c: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + Selfdeclare +
## lmer.clas5.c:     PachListen + PianoPlay + (Instrument + Harmony | Subject)
## lmer.clas5.d: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas5.d:     Selfdeclare + PachListen + ClsListen + (Instrument + Harmony |
## lmer.clas5.d:     Subject)
##               Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## lmer.clas5sub 36 6190.1 6382.3 -3059.0   6118.1
## lmer.clas5.a  39 6184.7 6393.0 -3053.3   6106.7 11.3758      3  0.0098579
## lmer.clas5.b  44 6189.9 6424.9 -3051.0   6101.9  4.7549      5  0.4465243
## lmer.clas5.c  47 6189.2 6440.1 -3047.6   6095.2  6.7789      3  0.0792889
## lmer.clas5.d  51 6178.2 6450.6 -3038.1   6076.2 18.9132      4  0.0008174
##
## lmer.clas5sub
## lmer.clas5.a  **
## lmer.clas5.b
## lmer.clas5.c  .
## lmer.clas5.d  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F test suggest that only the model with only PianoPlay added is preferred

```r
lmer.clas5.f <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice  + PianoPlay  + ClsLis
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas5.a, lmer.clas5.f)
```

```
## Data: music.sub
## Models:
## lmer.clas5.a: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas5.a:     (Instrument + Harmony | Subject)
## lmer.clas5.f: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas5.f:     ClsListen + (Instrument + Harmony | Subject)
##              Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.clas5.a 39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas5.f 43 6187.9 6417.5 -3051.0   6101.9 4.768      4     0.3119
```

testing for random slope on PianoPlay. All three approaches suggest not adding random slope on PianoPlay

```r
lmer.clas6<- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice  +PianoPlay +  (Instrumen
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas6.a <- lmer(Classical ~ Instrument + Harmony  +  Voice +  Harmony:Voice  +PianoPlay +  (Instru
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : unable to evaluate scaled gradient

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge: degenerate Hessian with 1
## negative eigenvalues
```

```r
lmer.clas6.formatted <- lmer(Classical ~Instrument + Harmony  +  Voice +  Harmony:Voice  +PianoPlay + (
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0296846
## (tol = 0.002, component 1)
```

lmer.clas6.aic <- fitLMER.fnc(lmer.clas6.formatted, ran.effects = c("(PianoPlay | Subject)"), method = "AIC") lmer.clas6.t <- fitLMER.fnc(lmer.clas6.formatted, ran.effects = c("(PianoPlay | Subject)"), method = "t") lmer.clas6.llrt <- fitLMER.fnc(lmer.clas6.formatted, ran.effects = c("(PianoPlay | Subject)"), method = "llrt")

```r
anova(lmer.clas6.a, lmer.clas6.aic, lmer.clas6.t, lmer.clas6.llrt)
```

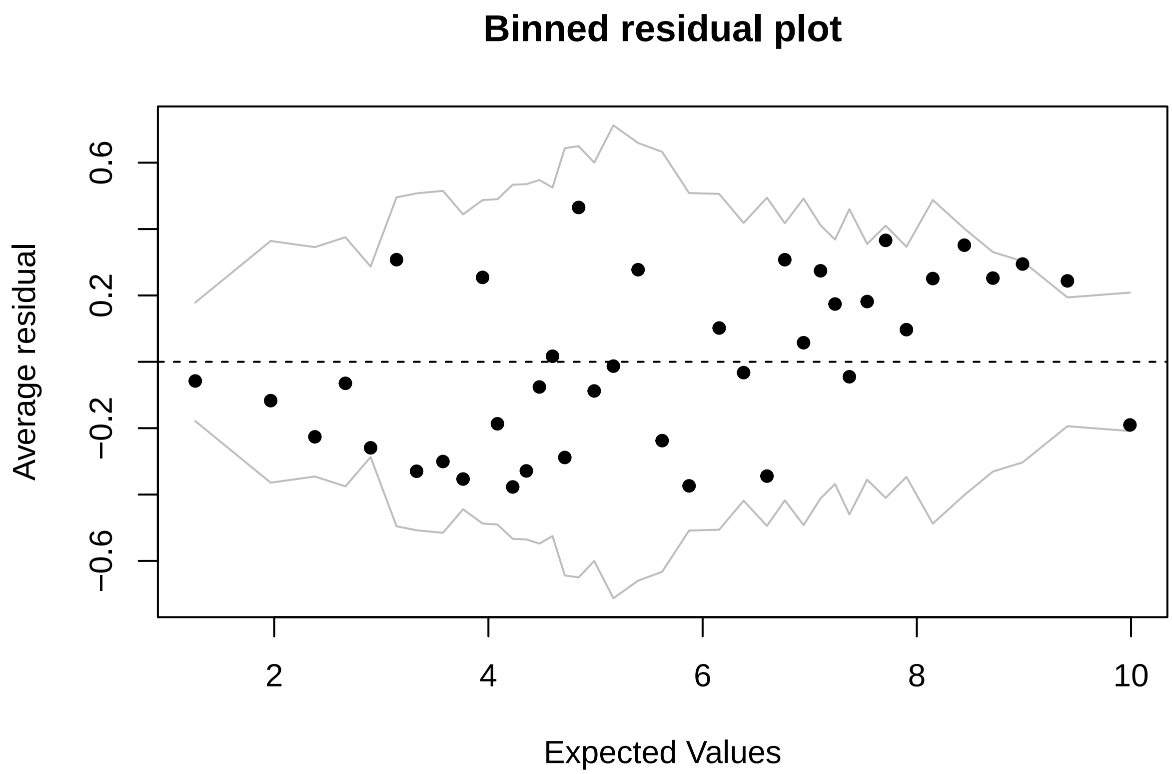
```
## refitting model(s) with ML (instead of REML)

## Data: music.sub
## Models:
## lmer.clas6.aic: Classical ~ Instrument + Harmony + Voice + PianoPlay + (1 + Instrument |
## lmer.clas6.aic:     Subject) + (0 + Harmony | Subject) + Harmony:Voice
## lmer.clas6.t: Classical ~ Instrument + Harmony + Voice + PianoPlay + (1 + Instrument |
## lmer.clas6.t:     Subject) + (0 + Harmony | Subject) + Harmony:Voice
## lmer.clas6.llrt: Classical ~ Instrument + Harmony + Voice + PianoPlay + (1 + Instrument |
## lmer.clas6.llrt:     Subject) + (0 + Harmony | Subject) + Harmony:Voice
## lmer.clas6.a: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas6.a:     (Instrument + Harmony + PianoPlay | Subject)
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.clas6.aic  34 6195.0 6376.6 -3063.5   6127.0
## lmer.clas6.t    34 6195.0 6376.6 -3063.5   6127.0  0.000      0    1.00000
## lmer.clas6.llrt 34 6195.0 6376.6 -3063.5   6127.0  0.000      0    1.00000
## lmer.clas6.a    63 6206.2 6542.6 -3040.1   6080.2 46.821     29    0.01941
##
## lmer.clas6.aic
## lmer.clas6.t
## lmer.clas6.llrt
## lmer.clas6.a    *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lmer.clas6)
```

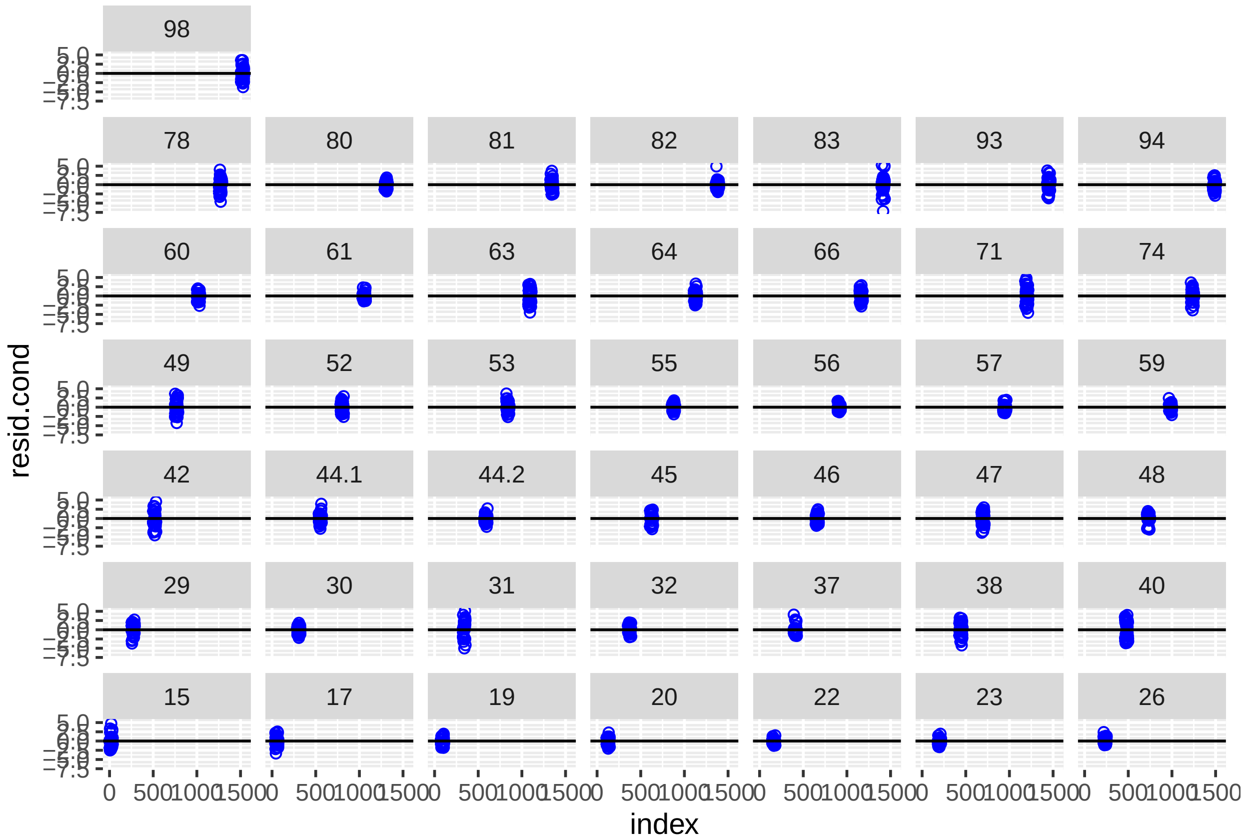
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
##     (Instrument + Harmony | Subject)
##    Data: music.sub
##
##      AIC      BIC   logLik deviance df.resid
##   6184.7   6393.0  -3053.3   6106.7     1502
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6114 -0.5687  0.0255  0.5153  3.4009
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      1.3277   1.1523
##           Instrumentpiano  1.9261   1.3879   -0.27
##           Instrumentstring 3.6838   1.9193   -0.55  0.62
##           HarmonyI-V-IV    0.1356   0.3682    0.85 -0.57 -0.69
##           HarmonyI-V-VI    1.7645   1.3284    0.15 -0.39 -0.59  0.45
##           HarmonyIV-I-V    0.1405   0.3748    0.27 -0.24 -0.26  0.08  0.41
##  Residual                  2.4321   1.5595
## Number of obs: 1541, groups:  Subject, 43
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                  3.3681     0.2716  12.402
## Instrumentpiano              1.6526     0.2331   7.090
## Instrumentstring             3.5878     0.3084  11.634
## HarmonyI-V-IV                0.2113     0.2026   1.043
## HarmonyI-V-VI                1.2657     0.2812   4.500
## HarmonyIV-I-V               -0.3023     0.2024  -1.494
## Voicepar3rd                 -0.3101     0.1942  -1.597
## Voicepar5th                 -0.2033     0.1946  -1.045
## PianoPlay1                   0.6110     0.3620   1.688
## PianoPlay4                   1.9854     0.6237   3.183
## PianoPlay5                   1.0874     0.5013   2.169
## HarmonyI-V-IV:Voicepar3rd  -0.4284     0.2749  -1.558
## HarmonyI-V-VI:Voicepar3rd  -0.7072     0.2755  -2.567
## HarmonyIV-I-V:Voicepar3rd   0.7515     0.2749   2.733
## HarmonyI-V-IV:Voicepar5th  -0.2107     0.2755  -0.765
## HarmonyI-V-VI:Voicepar5th  -0.5236     0.2756  -1.900
## HarmonyIV-I-V:Voicepar5th   0.3351     0.2749   1.219

##
## Correlation matrix not shown by default, as p = 17 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Checking condition residual plots. The binned plot shows all residual inside the 95% CI, which is preferred. The facet plot shows all residuals centered at 0, which is preferred. The Q-Q plot displays issue of long tail on both ends, suggesting maybe further transformation should be specified.

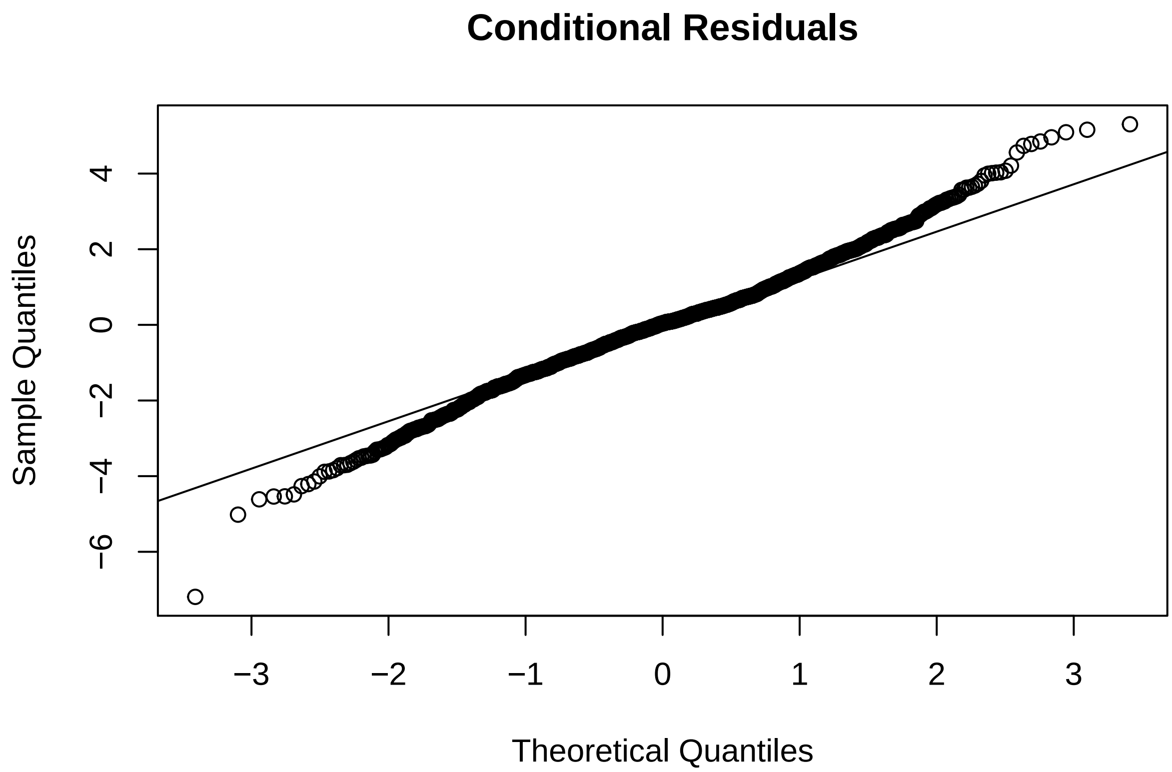
```
binnedplot(fit.cond,resid.cond)
```




```
index <- 1:dim(music.sub)[1]

new.data <- data.frame(index,resid.cond,music.sub$Subject)
names(new.data) <- c("index","resid.cond","Subject")
ggplot(new.data,aes(x=index,y=resid.cond)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```




```r
qqnorm(resid.cond,main="Conditional Residuals")
qqline(resid.cond)
```



Testing covariates for Popular

```r
lm.pop.fixed_only <- lm(Popular ~ Instrument + Harmony * Voice  +  Selfdeclare + log_OMSI  +Instr.minus
lm.pop.fixedonly.trim <- stepAIC(lm.pop.fixed_only,direction="both",trace=0)
summary(lm.pop.fixedonly.trim)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony + Voice + Selfdeclare +
##     log_OMSI + Instr.minus.Notes + PachListen + ClsListen + CollegeMusic +
##     KnowRob + KnowAxis + NoClass + PianoPlay + Composing + Harmony:Voice,
##     data = music.sub)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.990 -1.336  0.006  1.349 10.216
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.34422    0.59068  10.741  < 2e-16 ***
## Instrumentpiano            -1.14207    0.12815  -8.912  < 2e-16 ***
## Instrumentstring           -3.01924    0.12773 -23.637  < 2e-16 ***
## HarmonyI-V-IV               0.04601    0.25584   0.180 0.857299
```

```
## HarmonyI-V-VI            -0.19673    0.25635  -0.767 0.442948
## HarmonyIV-I-V             0.34884    0.25534   1.366 0.172083
## Voicepar3rd               0.48062    0.25534   1.882 0.059987 .
## Voicepar5th               0.46789    0.25584   1.829 0.067627 .
## Selfdeclare2              0.47234    0.25301   1.867 0.062108 .
## Selfdeclare3             -0.67283    0.28471  -2.363 0.018242 *
## Selfdeclare4             -0.78629    0.44200  -1.779 0.075449 .
## Selfdeclare5              0.67153    0.60552   1.109 0.267602
## Selfdeclare6             -4.56733    1.07708  -4.240 2.37e-05 ***
## log_OMSI                  0.31642    0.10063   3.144 0.001697 **
## Instr.minus.Notes        -0.13569    0.05045  -2.689 0.007237 **
## PachListen3              -2.17988    0.40875  -5.333 1.11e-07 ***
## PachListen4              -2.35940    0.63524  -3.714 0.000211 ***
## PachListen5              -2.92009    0.35343  -8.262 3.10e-16 ***
## ClsListen1                1.91644    0.22171   8.644  < 2e-16 ***
## ClsListen3                1.00172    0.21152   4.736 2.39e-06 ***
## ClsListen4                1.49925    0.77625   1.931 0.053620 .
## ClsListen5                0.71618    0.26101   2.744 0.006144 **
## CollegeMusic1             0.31215    0.20175   1.547 0.122035
## KnowRob1                  0.59759    0.28967   2.063 0.039279 *
## KnowRob5                  1.84006    0.27440   6.706 2.83e-11 ***
## KnowAxis1                 0.92536    0.52384   1.767 0.077512 .
## KnowAxis5                -0.52647    0.19720  -2.670 0.007674 **
## NoClass                   0.40496    0.11318   3.578 0.000357 ***
## PianoPlay1               -1.95209    0.25960  -7.519 9.41e-14 ***
## PianoPlay4               -0.08740    0.24470  -0.357 0.721017
## PianoPlay5               -0.45529    0.33180  -1.372 0.170209
## Composing                 0.21065    0.06320   3.333 0.000881 ***
## HarmonyI-V-IV:Voicepar3rd -0.01500    0.36146  -0.042 0.966895
## HarmonyI-V-VI:Voicepar3rd -0.10506    0.36217  -0.290 0.771783
## HarmonyIV-I-V:Voicepar3rd -1.00220    0.36146  -2.773 0.005629 **
## HarmonyI-V-IV:Voicepar5th -0.05382    0.36216  -0.149 0.881874
## HarmonyI-V-VI:Voicepar5th -0.07736    0.36219  -0.214 0.830905
## HarmonyIV-I-V:Voicepar5th -0.79347    0.36146  -2.195 0.028302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.051 on 1503 degrees of freedom
## Multiple R-squared:  0.3892, Adjusted R-squared:  0.3742
## F-statistic: 25.89 on 37 and 1503 DF,  p-value: < 2.2e-16
```

Adding the candidate fixed effects in lmer and use fitLMER.fnc() to see which ones were kept. PachListen was suggested by all three approaches. However after adding PachListen to model, the lmer model failed to converge. Hence will not add it into the model, and lmer.pop2 remains as our best model

```
lmer.pop.ran <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare +
    log_OMSI + Instr.minus.Notes + PachListen + ClsListen + CollegeMusic +
    KnowRob + KnowAxis + NoClass + PianoPlay + Composing + (Instrument  | Subject),  data = music.sub, 
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0964605
## (tol = 0.002, component 1)
```

lmer.pop.ran.aic <- fitLMER.fnc(lmer.pop.ran,method = “AIC”) lmer.pop.ran.t <- fitLMER.fnc(lmer.pop.ran,method = “t”) lmer.pop.ran.llrt <- fitLMER.fnc(lmer.pop.ran,method = “llrt”)

```
anova(lmer.pop.ran.aic,lmer.pop.ran.t,lmer.pop.ran.llrt)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: music.sub
## Models:
## lmer.pop.ran.aic: Popular ~ Instrument + PachListen + (Instrument | Subject)
## lmer.pop.ran.llrt: Popular ~ Instrument + Harmony + PachListen + (Instrument | Subject)
## lmer.pop.ran.t: Popular ~ Instrument + Harmony + PachListen + KnowRob + (Instrument |
## lmer.pop.ran.t:     Subject)
##                   Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop.ran.aic  13 6356.7 6426.1 -3165.4   6330.7
## lmer.pop.ran.llrt 16 6354.0 6439.5 -3161.0   6322.0 8.6722      3
## lmer.pop.ran.t    18 6352.3 6448.4 -3158.2   6316.3 5.7408      2
##                   Pr(>Chisq)
## lmer.pop.ran.aic
## lmer.pop.ran.llrt    0.03398 *
## lmer.pop.ran.t       0.05668 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

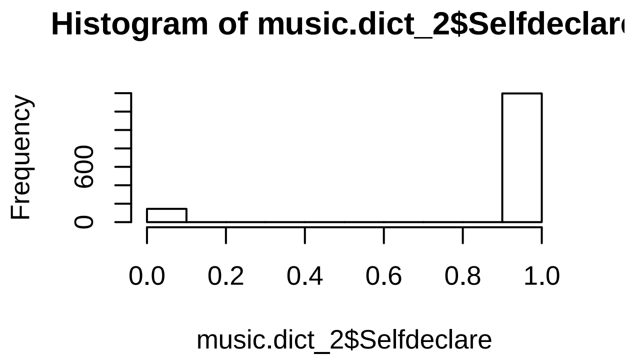
```
lmer.pop.pach <- lmer(Popular ~ Instrument + Harmony + Voice + PachListen + (Instrument|Subject),data =
```

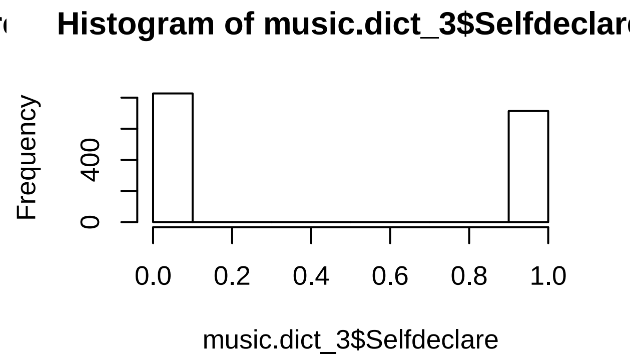
```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00217306
## (tol = 0.002, component 1)
```

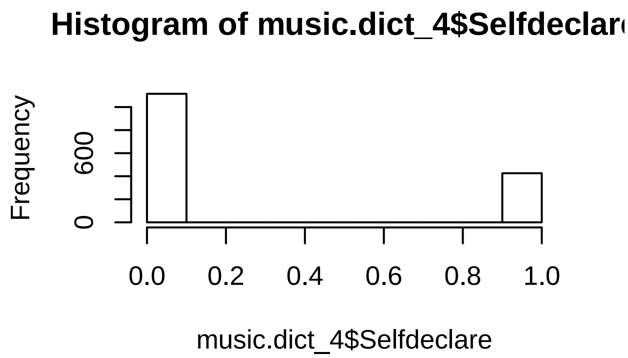
Testing self-identification as musician for Classical

Finding best cut-off to dichotomize Selfdeclare, seems like 3 is the best cutoff that gives roughly 50-50 split

```
music.dict_2 <- music.sub %>% mutate(Selfdeclare = ifelse(Selfdeclare %in% c(2,3,4,5,6),1, 0))
music.dict_3 <- music.sub %>% mutate(Selfdeclare = ifelse(Selfdeclare %in% c(3,4,5,6), 1, 0))
music.dict_4  <- music.sub %>% mutate(Selfdeclare = ifelse(Selfdeclare %in% c(4,5,6), 1, 0))
par(mfrow=c(2,2))
hist(music.dict_2$Selfdeclare)
hist(music.dict_3$Selfdeclare)
hist(music.dict_4$Selfdeclare)
```







Checking whether the selfdeclare should be added as fixed effect

```
lmer.clas73 <- lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
   (Instrument + Harmony | Subject),data = music.dict_3,REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```
lmer.clas6.dict3 <-  lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +  Selfde
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.134677
## (tol = 0.002, component 1)
```

```
anova(lmer.clas73,lmer.clas6.dict3)
```

```
## Data: music.dict_3
## Models:
## lmer.clas73: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas73:     (Instrument + Harmony | Subject)
## lmer.clas6.dict3: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas6.dict3:     Selfdeclare + (Instrument + Harmony | Subject)
##                  Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.clas73      39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas6.dict3 40 6186.7 6400.4 -3053.4   6106.7     0      1          1
```

Checking whether interaction between selfdeclare and the existing variables are significant. Found that none of the interaction terms are important

```r
lmer.clas73.intins<- update(lmer.clas6.dict3, .~.+Instrument*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas73.intharm<- update(lmer.clas6.dict3, .~.+Harmony*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas73.intvoi<- update(lmer.clas6.dict3, .~.+Voice*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas73.intpia<- update(lmer.clas6.dict3, .~.+PianoPlay*Selfdeclare)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas73,lmer.clas73.intins)
```

```
## Data: music.dict_3
## Models:
## lmer.clas73: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas73:     (Instrument + Harmony | Subject)
## lmer.clas73.intins: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas73.intins:     (Instrument + Harmony | Subject) + Harmony:Voice + Instrument:Selfdeclare
##                    Df    AIC    BIC  logLik deviance Chisq Chi Df
## lmer.clas73        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas73.intins 42 6190.4 6414.7 -3053.2   6106.4 0.321      3
##                    Pr(>Chisq)
## lmer.clas73
## lmer.clas73.intins      0.956
```

```r
anova(lmer.clas73,lmer.clas73.intharm)
```

```
## Data: music.dict_3
## Models:
## lmer.clas73: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas73:     (Instrument + Harmony | Subject)
## lmer.clas73.intharm: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas73.intharm:     (Instrument + Harmony | Subject) + Harmony:Voice + Harmony:Selfdeclare
##                     Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas73         39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas73.intharm 43 6186.4 6416.1 -3050.2   6100.4 6.2524      4
##                     Pr(>Chisq)
## lmer.clas73
## lmer.clas73.intharm     0.1811
```

```r
anova(lmer.clas73,lmer.clas73.intvoi)
```

```
## Data: music.dict_3
## Models:
## lmer.clas73: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas73:     (Instrument + Harmony | Subject)
## lmer.clas73.intvoi: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas73.intvoi:     (Instrument + Harmony | Subject) + Harmony:Voice + Voice:Selfdeclare
##                    Df    AIC  BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.clas73        39 6184.7 6393 -3053.3   6106.7
## lmer.clas73.intvoi 42 6189.7 6414 -3052.9   6105.7  0.94      3     0.8158
```

```r
anova(lmer.clas73,lmer.clas73.intpia)
```

```
## Data: music.dict_3
## Models:
## lmer.clas73: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas73:     (Instrument + Harmony | Subject)
## lmer.clas73.intpia: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas73.intpia:     (Instrument + Harmony | Subject) + Harmony:Voice + PianoPlay:Selfdeclare
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas73        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas73.intpia 42 6187.4 6411.7 -3051.7   6103.4 3.3154      3
##                    Pr(>Chisq)
## lmer.clas73
## lmer.clas73.intpia     0.3455
```

Test sensitivity to dichotomization. Dichotomizing at 2 or 4 still suggest that interaction between Selfdeclare and the three main variables should not be added.

```r
lmer.clas72 <- lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
   (Instrument + Harmony | Subject),data = music.dict_2,REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas6.dict2 <-  lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +  Selfdec
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas72,lmer.clas6.dict2)
```

```
## Data: music.dict_2
## Models:
## lmer.clas72: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas72:     (Instrument + Harmony | Subject)
## lmer.clas6.dict2: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas6.dict2:     Selfdeclare + (Instrument + Harmony | Subject)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas72      39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas6.dict2 40 6186.3 6399.9 -3053.1   6106.3 0.4364      1
##                  Pr(>Chisq)
## lmer.clas72
## lmer.clas6.dict2     0.5089
```

```r
lmer.clas72.intins<- update(lmer.clas6.dict2, .~.+Instrument*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas72.intharm<- update(lmer.clas6.dict2, .~.+Harmony*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas72.intvoi<- update(lmer.clas6.dict2, .~.+Voice*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas72.intpia<- update(lmer.clas6.dict2, .~.+PianoPlay*Selfdeclare)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas72,lmer.clas72.intins)
```

```
## Data: music.dict_2
## Models:
## lmer.clas72: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas72:     (Instrument + Harmony | Subject)
## lmer.clas72.intins: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas72.intins:     (Instrument + Harmony | Subject) + Harmony:Voice + Instrument:Selfdeclare
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas72        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas72.intins 42 6186.0 6410.3 -3051.0   6102.0 4.7033      3
##                    Pr(>Chisq)
## lmer.clas72
## lmer.clas72.intins     0.1949
```

```r
anova(lmer.clas72,lmer.clas72.intharm)
```

```
## Data: music.dict_2
## Models:
## lmer.clas72: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas72:     (Instrument + Harmony | Subject)
## lmer.clas72.intharm: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas72.intharm:     (Instrument + Harmony | Subject) + Harmony:Voice + Harmony:Selfdeclare
##                     Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas72         39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas72.intharm 43 6184.1 6413.7 -3049.0   6098.1 8.6093      4
##                     Pr(>Chisq)
## lmer.clas72
## lmer.clas72.intharm    0.07164 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.clas72,lmer.clas72.intvoi)
```

```
## Data: music.dict_2
## Models:
## lmer.clas72: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas72:     (Instrument + Harmony | Subject)
## lmer.clas72.intvoi: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas72.intvoi:     (Instrument + Harmony | Subject) + Harmony:Voice + Voice:Selfdeclare
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas72        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas72.intvoi 42 6187.8 6412.1 -3051.9   6103.8 2.8614      3
##                    Pr(>Chisq)
## lmer.clas72
## lmer.clas72.intvoi     0.4135
```

```r
anova(lmer.clas72,lmer.clas72.intpia)
```

```
## Data: music.dict_2
## Models:
## lmer.clas72: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas72:     (Instrument + Harmony | Subject)
## lmer.clas72.intpia: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas72.intpia:     (Instrument + Harmony | Subject) + Harmony:Voice + PianoPlay:Selfdeclare
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas72        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas72.intpia 42 6186.5 6410.8 -3051.3   6102.5 4.1518      3
##                    Pr(>Chisq)
## lmer.clas72
## lmer.clas72.intpia     0.2455
```

```r
lmer.clas74 <- lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
   (Instrument + Harmony | Subject),data = music.dict_4,REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas6.dict4 <-  lmer(Classical~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +  Selfde
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas74,lmer.clas6.dict4)
```

```
## Data: music.dict_4
## Models:
## lmer.clas74: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas74:     (Instrument + Harmony | Subject)
## lmer.clas6.dict4: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas6.dict4:     Selfdeclare + (Instrument + Harmony | Subject)
##                  Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.clas74      39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas6.dict4 40 6187.3 6400.9 -3053.7   6107.3     0      1          1
```

```r
lmer.clas74.intins<- update(lmer.clas6.dict4, .~.+Instrument*Selfdeclare)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0222027
## (tol = 0.002, component 1)
```

```r
lmer.clas74.intharm<- update(lmer.clas6.dict4, .~.+Harmony*Selfdeclare)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0309593
## (tol = 0.002, component 1)
```

```r
lmer.clas74.intvoi<- update(lmer.clas6.dict4, .~.+Voice*Selfdeclare)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
lmer.clas74.intpia<- update(lmer.clas6.dict4, .~.+PianoPlay*Selfdeclare)
```

```
## fixed-effect model matrix is rank deficient so dropping 2 columns / coefficients
## boundary (singular) fit: see ?isSingular
```

```r
anova(lmer.clas74,lmer.clas74.intins)
```

```
## Data: music.dict_4
## Models:
## lmer.clas74: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas74:     (Instrument + Harmony | Subject)
## lmer.clas74.intins: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas74.intins:     (Instrument + Harmony | Subject) + Harmony:Voice + Instrument:Selfdeclare
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas74        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas74.intins 42 6189.8 6414.1 -3052.9   6105.8 0.8827      3
##                    Pr(>Chisq)
## lmer.clas74
## lmer.clas74.intins     0.8296
```

```r
anova(lmer.clas74,lmer.clas74.intharm)
```

```
## Data: music.dict_4
## Models:
## lmer.clas74: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas74:     (Instrument + Harmony | Subject)
## lmer.clas74.intharm: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas74.intharm:     (Instrument + Harmony | Subject) + Harmony:Voice + Harmony:Selfdeclare
##                     Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.clas74         39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas74.intharm 43 6185.7 6415.3 -3049.8   6099.7 7.0045      4
##                     Pr(>Chisq)
## lmer.clas74
## lmer.clas74.intharm     0.1356
```

```r
anova(lmer.clas74,lmer.clas74.intvoi)
```

```
## Data: music.dict_4
## Models:
## lmer.clas74: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas74:     (Instrument + Harmony | Subject)
## lmer.clas74.intvoi: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas74.intvoi:     (Instrument + Harmony | Subject) + Harmony:Voice + Voice:Selfdeclare
##                    Df    AIC    BIC  logLik deviance Chisq Chi Df
## lmer.clas74        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas74.intvoi 42 6191.1 6415.4 -3053.6   6107.1     0      3
##                    Pr(>Chisq)
## lmer.clas74
## lmer.clas74.intvoi          1
```

```r
anova(lmer.clas74,lmer.clas74.intpia)
```

```
## Data: music.dict_4
## Models:
## lmer.clas74: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + PianoPlay +
## lmer.clas74:     (Instrument + Harmony | Subject)
## lmer.clas74.intpia: Classical ~ Instrument + Harmony + Voice + PianoPlay + Selfdeclare +
## lmer.clas74.intpia:     (Instrument + Harmony | Subject) + Harmony:Voice + PianoPlay:Selfdeclare
##                    Df    AIC    BIC  logLik deviance Chisq Chi Df
## lmer.clas74        39 6184.7 6393.0 -3053.3   6106.7
## lmer.clas74.intpia 41 6189.1 6408.1 -3053.6   6107.1     0      2
##                    Pr(>Chisq)
## lmer.clas74
## lmer.clas74.intpia          1
```

Testing self-identification as musician for Popular

Checking whether the selfdeclare should be added as fixed effect

```r
lmer.pop2.dict3 <- lmer( Popular ~ Instrument + Harmony + Voice +(1+ Instrument | Subject), data = musi

lmer.pop3.dict3 <- lmer( Popular ~ Instrument + Harmony + Voice + Selfdeclare +  (1 + Instrument | Subj

anova(lmer.pop2.dict3,lmer.pop3.dict3)
```

```
## Data: music.dict_3
## Models:
## lmer.pop2.dict3: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop3.dict3: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict3:     Subject)
##                 Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.pop2.dict3 15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict3 16 6360.5 6445.9 -3164.2   6328.5 3e-04      1      0.986
```

Checking whether interaction between selfdeclare and the existing variables are significant. Found that the interaction term Harmony*Selfdeclare is important. Hence lmer.pop.intharm3, which is Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument | Subject) + Harmony:Selfdeclare is our best model here.

```r
lmer.pop.intins3<- update(lmer.pop3.dict3 , .~.+Instrument*Selfdeclare)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00308461
## (tol = 0.002, component 1)
```

```r
lmer.pop.intharm3<- update(lmer.pop3.dict3 , .~.+Harmony*Selfdeclare)
lmer.pop.intvoi3<- update(lmer.pop3.dict3 , .~.+Voice*Selfdeclare)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0242068
## (tol = 0.002, component 1)
```

```r
anova(lmer.pop2.dict3 ,lmer.pop.intins3)
```

```
## Data: music.dict_3
## Models:
## lmer.pop2.dict3: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intins3: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intins3:     Subject) + Instrument:Selfdeclare
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict3  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intins3 18 6359.9 6456.1 -3162.0   6323.9 4.5187      3
##                  Pr(>Chisq)
## lmer.pop2.dict3
## lmer.pop.intins3     0.2106
```

```r
anova(lmer.pop2.dict3 ,lmer.pop3.dict3, lmer.pop.intharm3)
```

```
## Data: music.dict_3
## Models:
## lmer.pop2.dict3: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop3.dict3: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict3:     Subject)
## lmer.pop.intharm3: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intharm3:     Subject) + Harmony:Selfdeclare
##                   Df    AIC    BIC  logLik deviance   Chisq Chi Df
## lmer.pop2.dict3   15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict3   16 6360.5 6445.9 -3164.2   6328.5  0.0003      1
## lmer.pop.intharm3 19 6350.2 6451.6 -3156.1   6312.2 16.2953      3
##                   Pr(>Chisq)
## lmer.pop2.dict3
## lmer.pop3.dict3    0.9860067
## lmer.pop.intharm3  0.0009863 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.pop2.dict3 ,lmer.pop.intvoi3)
```

```
## Data: music.dict_3
## Models:
## lmer.pop2.dict3: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intvoi3: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intvoi3:     Subject) + Voice:Selfdeclare
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict3  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intvoi3 18 6362.6 6458.7 -3163.3   6326.6 1.8551      3
##                  Pr(>Chisq)
## lmer.pop2.dict3
## lmer.pop.intvoi3      0.603
```

```
summary(lmer.pop.intharm3)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
##     Subject) + Harmony:Selfdeclare
##    Data: music.dict_3
##
##      AIC      BIC   logLik deviance df.resid
##   6350.2   6451.6  -3156.1   6312.2     1522
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6564 -0.5856  0.0195  0.5969  5.2903
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      1.276    1.129
##           Instrumentpiano  1.727    1.314    -0.19
##           Instrumentstring 2.556    1.599    -0.36  0.72
##  Residual                  3.002    1.733
## Number of obs: 1541, groups:  Subject, 43
##
## Fixed effects:
##                              Estimate Std. Error t value
## (Intercept)                 6.748e+00  2.754e-01  24.500
## Instrumentpiano            -1.148e+00  2.279e-01  -5.039
## Instrumentstring           -3.024e+00  2.666e-01 -11.341
## HarmonyI-V-IV              -7.268e-14  1.703e-01   0.000
## HarmonyI-V-VI               1.318e-01  1.705e-01   0.773
## HarmonyIV-I-V              -2.174e-01  1.703e-01  -1.276
## Voicepar3rd                 1.953e-01  1.081e-01   1.806
## Voicepar5th                 2.323e-01  1.081e-01   2.148
## Selfdeclare                 2.047e-01  3.764e-01   0.544
## HarmonyI-V-IV:Selfdeclare  5.054e-02  2.503e-01   0.202
## HarmonyI-V-VI:Selfdeclare -8.287e-01  2.505e-01  -3.308
## HarmonyIV-I-V:Selfdeclare -6.773e-02  2.501e-01  -0.271
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HrI-V-IV HrI-V-VI HrIV-I-V Vcpr3r
## Instrumntpn -0.197
## Instrmntstr -0.283  0.672
```

```
## HrmnyI-V-IV -0.309  0.000     0.000
## HrmnyI-V-VI -0.308 -0.001    -0.001     0.499
## HrmnyIV-I-V -0.309  0.000     0.000     0.500    0.499
## Voicepar3rd -0.197  0.000     0.000     0.000   -0.002    0.000
## Voicepar5th -0.197  0.000     0.000     0.000   -0.002    0.000    0.501
## Selfdeclare -0.635  0.000     0.000     0.226    0.226    0.226    0.000
## HrmI-V-IV:S  0.211  0.001     0.000    -0.680   -0.340   -0.340   -0.001
## HrmI-V-VI:S  0.210  0.001     0.000    -0.340   -0.681   -0.340    0.001
## HrmIV-I-V:S  0.211  0.000     0.000    -0.340   -0.340   -0.681    0.001
##             Vcpr5t Slfdcl HI-V-IV: HI-V-VI:
## Instrumntpn
## Instrmntstr
## HrmnyI-V-IV
## HrmnyI-V-VI
## HrmnyIV-I-V
## Voicepar3rd
## Voicepar5th
## Selfdeclare  0.001
## HrmI-V-IV:S -0.001 -0.332
## HrmI-V-VI:S -0.001 -0.332  0.499
## HrmIV-I-V:S -0.001 -0.332  0.500    0.499
```

Testing whether result is sensitive to dichotomizing at 2 or 4

```r
lmer.pop2.dict2 <- lmer( Popular ~ Instrument + Harmony + Voice +(1+ Instrument | Subject), data = musi

lmer.pop3.dict2 <- lmer( Popular ~ Instrument + Harmony + Voice + Selfdeclare +  (1 + Instrument | Subj

anova(lmer.pop2.dict2,lmer.pop3.dict2)
```

```
## Data: music.dict_2
## Models:
## lmer.pop2.dict2: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop3.dict2: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict2:     Subject)
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.pop2.dict2 15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict2 16 6358.7 6444.1 -3163.3   6326.7 1.7825      1     0.1818
```

```r
lmer.pop.intins2<- update(lmer.pop3.dict2 , .~.+Instrument*Selfdeclare)
lmer.pop.intharm2<- update(lmer.pop3.dict2 , .~.+Harmony*Selfdeclare)
lmer.pop.intvoi2<- update(lmer.pop3.dict2, .~.+Voice*Selfdeclare)




anova(lmer.pop2.dict2 ,lmer.pop.intins2)
```

```
## Data: music.dict_2
## Models:
## lmer.pop2.dict2: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intins2: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intins2:     Subject) + Instrument:Selfdeclare
```

```
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict2  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intins2 18 6362.4 6458.5 -3163.2   6326.4 2.0919      3
##                 Pr(>Chisq)
## lmer.pop2.dict2
## lmer.pop.intins2     0.5535
```

```
anova(lmer.pop2.dict2 ,lmer.pop3.dict2, lmer.pop.intharm2)
```

```
## Data: music.dict_2
## Models:
## lmer.pop2.dict2: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop3.dict2: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict2:     Subject)
## lmer.pop.intharm2: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intharm2:     Subject) + Harmony:Selfdeclare
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict2   15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict2   16 6358.7 6444.1 -3163.3   6326.7 1.7825      1
## lmer.pop.intharm2 19 6358.1 6459.6 -3160.1   6320.1 6.5278      3
##                  Pr(>Chisq)
## lmer.pop2.dict2
## lmer.pop3.dict2      0.18184
## lmer.pop.intharm2    0.08857 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lmer.pop2.dict2 ,lmer.pop.intvoi2)
```

```
## Data: music.dict_2
## Models:
## lmer.pop2.dict2: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intvoi2: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intvoi2:     Subject) + Voice:Selfdeclare
##                 Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.pop2.dict2  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intvoi2 18 6361.9 6458.1 -3163.0   6325.9 2.529      3     0.4701
```

```
lmer.pop2.dict4 <- lmer( Popular ~ Instrument + Harmony + Voice +(1+ Instrument | Subject), data = musi

lmer.pop3.dict4 <- lmer( Popular ~ Instrument + Harmony + Voice + Selfdeclare +  (1 + Instrument | Subj
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00521951
## (tol = 0.002, component 1)
```

```
anova(lmer.pop2.dict4,lmer.pop3.dict4)
```

```
## Data: music.dict_4
## Models:
## lmer.pop2.dict4: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
```

```
## lmer.pop3.dict4: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict4:     Subject)
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.pop2.dict4 15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict4 16 6360.4 6445.9 -3164.2   6328.4 0.0314      1     0.8594
```

```
lmer.pop.intins4<- update(lmer.pop3.dict4 , .~.+Instrument*Selfdeclare)
lmer.pop.intharm4<- update(lmer.pop3.dict4 , .~.+Harmony*Selfdeclare)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00282762
## (tol = 0.002, component 1)
```

```
lmer.pop.intvoi4<- update(lmer.pop3.dict4, .~.+Voice*Selfdeclare)



anova(lmer.pop2.dict4 ,lmer.pop.intins4)
```

```
## Data: music.dict_4
## Models:
## lmer.pop2.dict4: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intins4: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intins4:     Subject) + Instrument:Selfdeclare
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict4  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intins4 18 6362.9 6459.1 -3163.5   6326.9 1.5203      3
##                  Pr(>Chisq)
## lmer.pop2.dict4
## lmer.pop.intins4     0.6776
```

```
anova(lmer.pop2.dict4 ,lmer.pop3.dict4, lmer.pop.intharm4)
```

```
## Data: music.dict_4
## Models:
## lmer.pop2.dict4: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop3.dict4: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop3.dict4:     Subject)
## lmer.pop.intharm4: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intharm4:     Subject) + Harmony:Selfdeclare
##                   Df    AIC    BIC  logLik deviance   Chisq Chi Df
## lmer.pop2.dict4   15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop3.dict4   16 6360.4 6445.9 -3164.2   6328.4  0.0314      1
## lmer.pop.intharm4 19 6346.6 6448.0 -3154.3   6308.6 19.8515      3
##                   Pr(>Chisq)
## lmer.pop2.dict4
## lmer.pop3.dict4    0.8593697
## lmer.pop.intharm4  0.0001822 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.pop2.dict4 ,lmer.pop.intvoi4)
```

```
## Data: music.dict_4
## Models:
## lmer.pop2.dict4: Popular ~ Instrument + Harmony + Voice + (1 + Instrument | Subject)
## lmer.pop.intvoi4: Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
## lmer.pop.intvoi4:     Subject) + Voice:Selfdeclare
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## lmer.pop2.dict4  15 6358.5 6438.6 -3164.2   6328.5
## lmer.pop.intvoi4 18 6363.7 6459.8 -3163.8   6327.7 0.7905      3
##                  Pr(>Chisq)
## lmer.pop2.dict4
## lmer.pop.intvoi4     0.8517
```

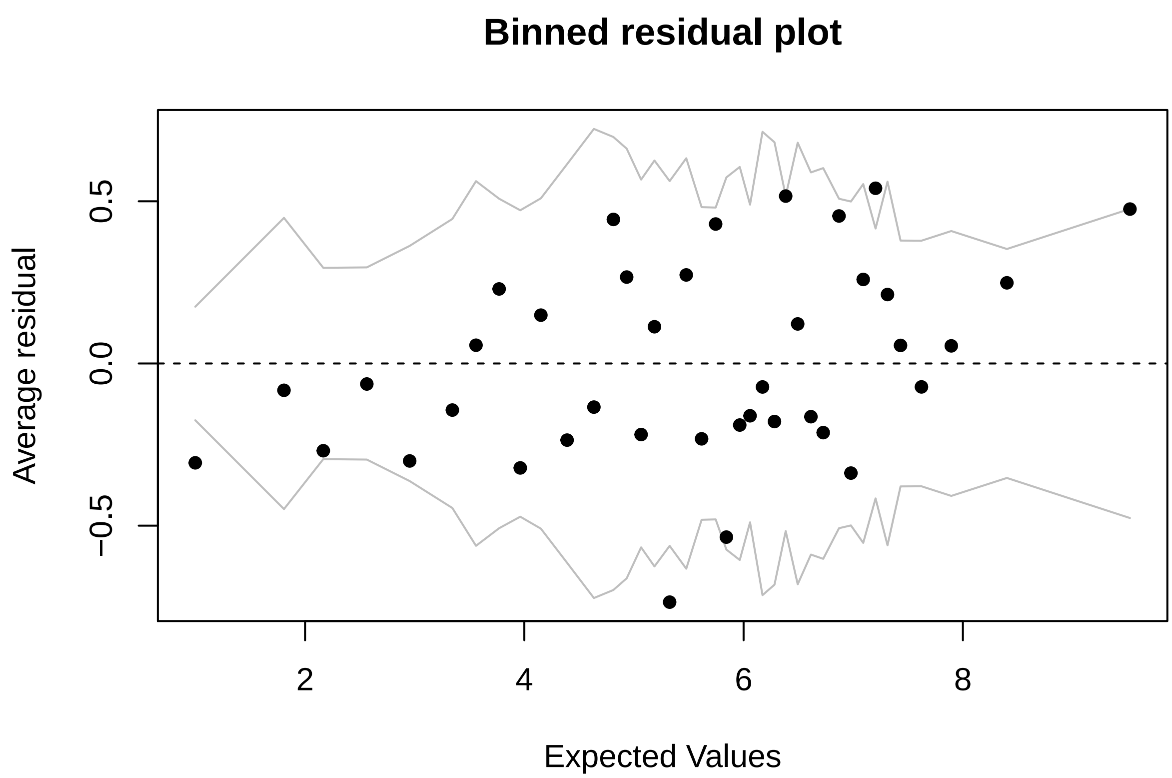
```r
summary(lmer.pop.intharm4)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Popular ~ Instrument + Harmony + Voice + Selfdeclare + (1 + Instrument |
##     Subject) + Harmony:Selfdeclare
##    Data: music.dict_4
##
##      AIC      BIC   logLik deviance df.resid
##   6346.6   6448.0  -3154.3   6308.6     1522
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5802 -0.5818  0.0177  0.5835  5.1950
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      1.271    1.127
##           Instrumentpiano  1.727    1.314    -0.19
##           Instrumentstring 2.557    1.599    -0.35  0.72
##  Residual                  2.995    1.731
## Number of obs: 1541, groups:  Subject, 43
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                 6.803971   0.242255  28.086
## Instrumentpiano            -1.148228   0.227812  -5.040
## Instrumentstring           -3.023636   0.266625 -11.340
## HarmonyI-V-IV              -0.039427   0.146519  -0.269
## HarmonyI-V-VI               0.007872   0.146660   0.054
## HarmonyIV-I-V              -0.215054   0.146519  -1.468
## Voicepar3rd                 0.195094   0.108005   1.806
## Voicepar5th                 0.232953   0.108005   2.157
## Selfdeclare                 0.139618   0.418508   0.334
## HarmonyI-V-IV:Selfdeclare   0.228391   0.278814   0.819
## HarmonyI-V-VI:Selfdeclare  -0.941906   0.278890  -3.377
## HarmonyIV-I-V:Selfdeclare  -0.121749   0.278290  -0.437
##
```

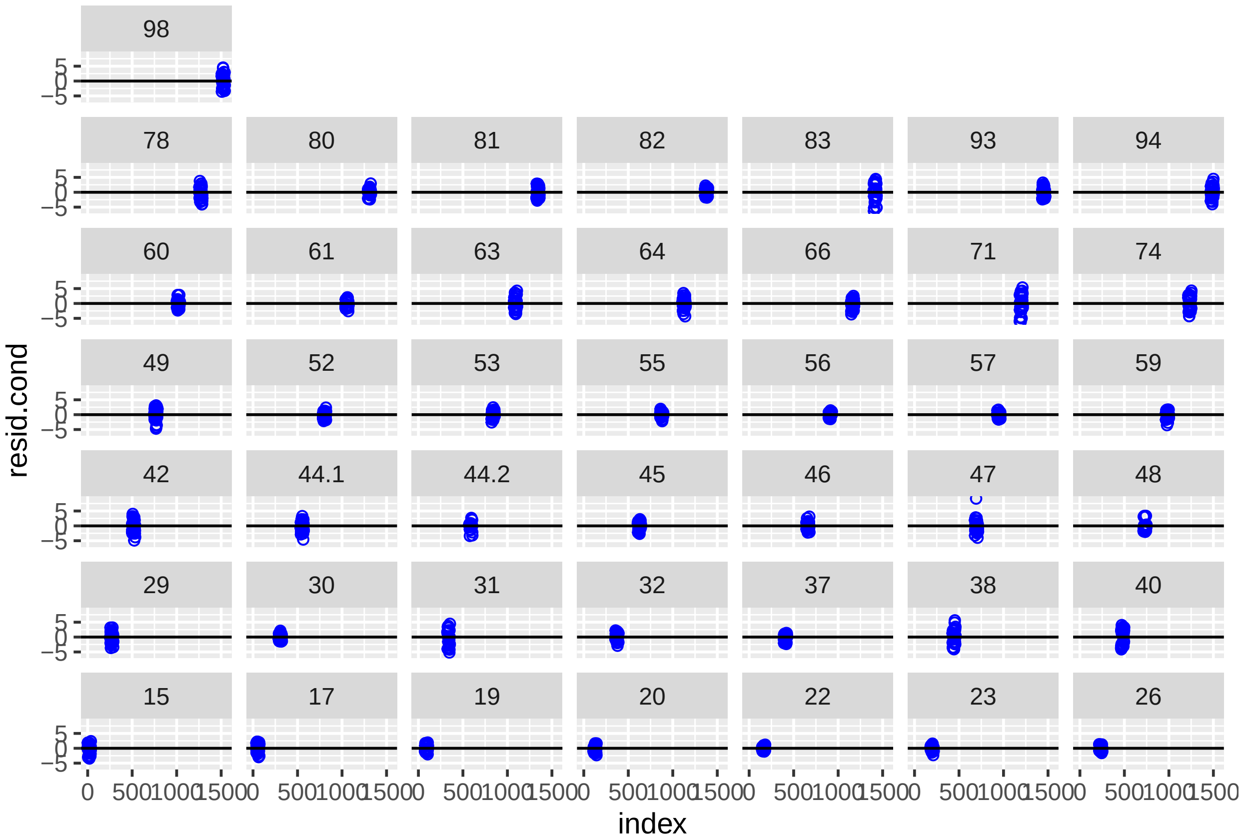
```
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HrI-V-IV HrI-V-VI HrIV-I-V Vcpr3r
## Instrumntpn -0.222
## Instrmntstr -0.319  0.672
## HrmnyI-V-IV -0.302  0.000     0.000
## HrmnyI-V-VI -0.301 -0.001    -0.001     0.500
## HrmnyIV-I-V -0.302  0.000     0.000     0.500    0.500
## Voicepar3rd -0.223  0.000     0.000     0.000   -0.001    0.000
## Voicepar5th -0.223  0.000     0.000     0.000   -0.001    0.000    0.501
## Selfdeclare -0.482  0.000     0.000     0.175    0.175    0.175    0.000
## HrmI-V-IV:S  0.159  0.001     0.000    -0.526   -0.262   -0.263   -0.002
## HrmI-V-VI:S  0.159  0.002     0.000    -0.263   -0.526   -0.263    0.001
## HrmIV-I-V:S  0.159  0.000     0.000    -0.263   -0.263   -0.526    0.002
##             Vcpr5t Slfdcl HI-V-IV: HI-V-VI:
## Instrumntpn
## Instrmntstr
## HrmnyI-V-IV
## HrmnyI-V-VI
## HrmnyIV-I-V
## Voicepar3rd
## Voicepar5th
## Selfdeclare  0.001
## HrmI-V-IV:S -0.002 -0.332
## HrmI-V-VI:S -0.003 -0.332  0.498
## HrmIV-I-V:S -0.002 -0.332  0.499    0.499
## convergence code: 0
## Model failed to converge with max|grad| = 0.00282762 (tol = 0.002, component 1)
```

Checking condition residual plots. The binned plot shows almost all residual inside the 95% CI, which is preferred. The facet plot shows all residuals centered at 0, which is preferred. The Q-Q plot displays a very slight long tail on both ends, but in general the issue is not too big.

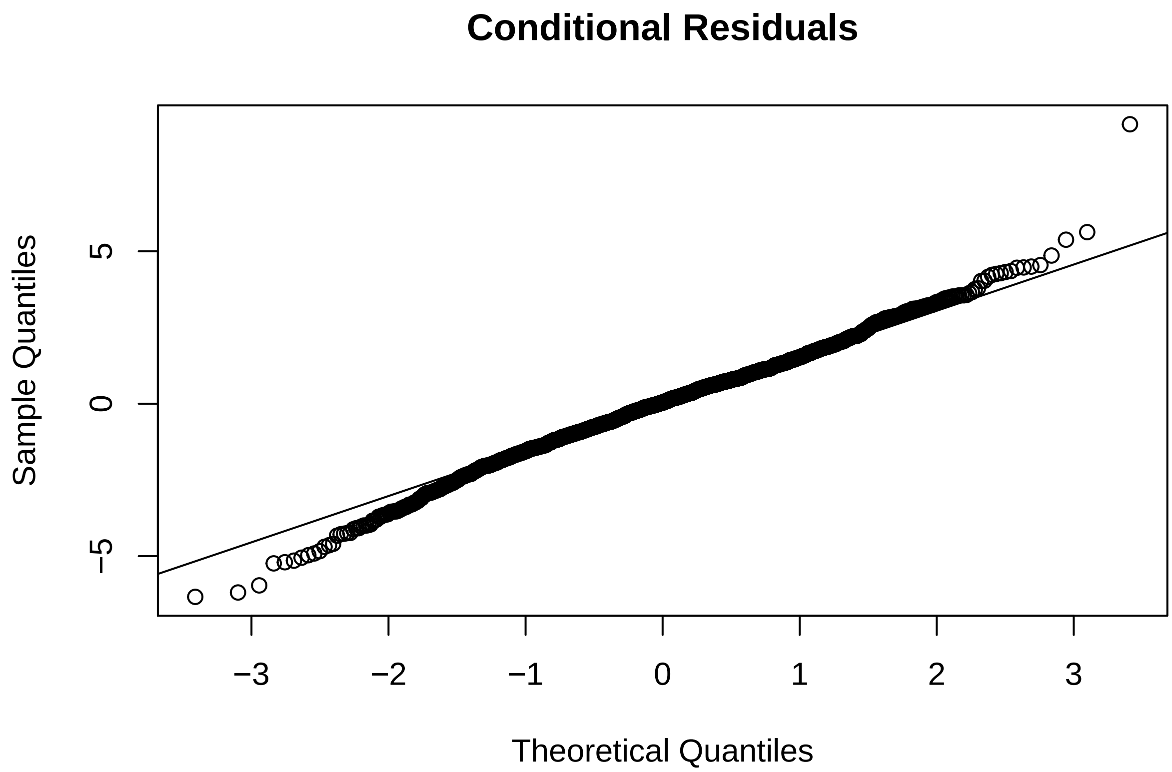
```r
binnedplot(fit.cond,resid.cond)
```




```
index <- 1:dim(music.dict_3)[1]

new.data <- data.frame(index,resid.cond,music.dict_3$Subject)
names(new.data) <- c("index","resid.cond","Subject")
ggplot(new.data,aes(x=index,y=resid.cond)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
    geom_hline(yintercept=0)
```




```
qqnorm(resid.cond,main="Conditional Residuals")
qqline(resid.cond)
```


Conditional Residuals
Sample Quantiles
5
0
−5
−3
−2
−1
0
1
2
3
Theoretical Quantiles