Examining Factors Leading to Identifying of Classical or Popular Music

Jae Won Yoon
jaewony@andrew.cmu.edu

December 8th, 2019

Abstract

We examine and measure the influence of instrument, harmonic motion and voice leading on listeners' identification and scoring of music as classical or popular. Based on the data collected from this experiment from the University of Pittsburgh, we discover that instrument exerts the strongest influence among the three design factors. We further discover that of the different harmonic motion levels, *I-V-VI* has the strongest association with classical ratings. Additionally, we determine that among the levels of voice leading, contrary motion has the strongest association with classical ratings. Furthermore, we discover that there are differences in ways musicians and non-musicians identify music. Finally, we conduct analysis on popular ratings and discover that instrument exerts the strongest influence. On the other hand, harmonic motion and voice leading is not significant to listeners' identification of popular music. For future analysis, this study should use data that does not contain a significant number of missing values.

1 Introduction

Researchers have often wondered what external factors influence listeners' classification of different types of music. In 2012, Ivan Jimenez, a composer and musicologist at the University of Pittsburgh, conducted a design experiment to examine the effects of instrument, harmonic motion and voice leading on listeners' identification of music as "classical" or "popular." Listeners would then decide rate how classical or how popular a music sounds on a scale of 1 to 10 (1 = not at all, 10 = very popular/classical sounding). Essentially, by examining the data collected from this experiment, we want to answer the following questions in our analysis:

- Determine which experimental factor, or combination of factors, has the strongest influence on ratings.

  - Determine if **Instrument** exerts the strongest influence among the three design factors (**Instrument, Harmonic Motion, Voice Leading**).
  - Determine if *I-V-VI* has a strong association with classical ratings among the levels of **Harmonic Motion**. Additionally, determine whether it seems to matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits.
  - Determine if contrary motion has a strong association with classical ratings among the levels of **Voice Leading**.

- Examine if there are differences in the way that musicians and non-musicians identify classical ratings.
- Identify if there are differences in the things that drive classical, vs. popular, ratings.

The experiment and the data was collected at the University of Pittsburgh in 2012. The analysis was conducted in R and steps are shown in the technical appendix.

2 Methods

The data set we use provides information on results collected from 70 University of Pittsburgh undergraduate students. They were presented with 36 musical stimuli and would then rate whether the song sounds classical or popular. The 36 stimuli were chosen by crossing these factors:

- **Instrument:** String Quartet, Piano, Electric Guitar
- **Harmonic Motion:** *I-V-VI*, I-VI-V, I-V-IV, IV-I-V
- **Voice Leading:** Contrary Motion, Parallel 3rds, Parallel 5ths

The data contains information on 26 different variables mainly related to instrumental, harmonic and vocal effects. There are two main response variables: classical and popular. In addition to the categorical variables representing each instrument, harmonic motion and voice leading, there are variables scored on different scales that represent answers to survey questions given to participants in the experiment. Most of these variables pertain to the three factors listed above. The full table that describes the variables in the data is listed below.

	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	Harmonic Motion	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 - X17	Auxiliary measure of listener's ability to distinguish classical vs popular music
10	ConsIntsr	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 - Notes	Difference between prev. two variables
13	PachListen	How familiar are you with Pachelbel's Canon in 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 - 1960s1970s	Difference between prev variable and a similar variable referring to 60's and 70's pop and rock.
19	CollegeMusic	Have you taken music classes in college (0=no, 1=yes)
20	NoClass	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	X2ndInstr	Same, for second musical instrument

Table 1: Variable Definitions of Data

The data set contains 2520 rows of data from the 36 participants in the study. To account for missing values in the data, we conduct median imputation to avoid deleting a large portion of the data. After processing the data, we conduct exploratory data analysis on classical and popular scores on the different levels of instrument, harmonic motion and voice leading.

To examine which experimental factor, or combination of factors, has the strongest influence on ratings, we first examine the influence of the three main experimental factors using conventional linear models and variable selection. We consider interactions between these three factors to account for the various levels of these variables. Additionally, since there are 36 ratings from each participant, we use analysis of covariance tests to test whether random effects are needed in our model. Afterwards, we add various combinations of random effects of the three experimental factors on the subjects to see which effects are most significant. After conducting multiple ANOVA tests, we select the model with lowest value of AIC to move forward.

Afterwards, we determine which individual covariates of the remaining 21 variables in the data should be added to the model as fixed effects. We conduct backward selection AIC to examine which predictors have significant effect on ratings. Once the fixed effects are settled, we go back and check to see whether there should be any changes in the random effects of our model. Our final will thus be a mixed model with fixed effects and random effects on subjects in the experiment.

To examine our second research question on if there are any differences in ways musicians and non-musicians identify classical music, we dichotomize the variable "Self-declare" so that about half the participants are categorized as self-declared musicians, and half not. We then conduct multiple ANOVA tests to see whether interactions between this dichotomized variable and the three factors are significant or not. Finally, we conduct sensitivity analysis on dichotomization with different methods of declaring whether one is a musician or not. Essentially, by altering the scale for musician and non-musician, we check to see if our results are sensitive to dichotomization.

Finally, to answer the last research question on whether there are differences between in what drives classical and popular ratings, we conduct an identical approach to popular ratings in that we use backward selection AIC on a random effects model with popular ratings as the response variable. We mainly check to see if there are any notable differences in the effect of instrument, harmonic motion (in particular, I-V-VI) and voice leading (in particular, contrary motion) on popular ratings.

All statistical methods conducted for this analysis are included in the technical appendix. The analysis is divided into 5 main sections. We will be referencing the technical appendix throughout the remainder of this report.

3 Results

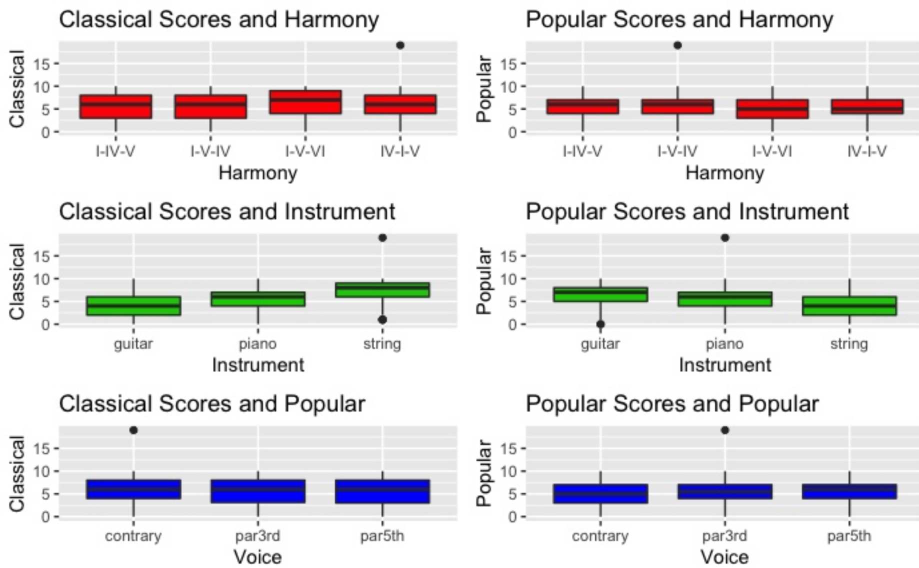


Figure 1: Classical and Popular Scores against the Three Experimental Factors

3.1 Exploratory Data Analysis

By examining box plots of classical and popular scores against instrument, harmony and voice in Figure 1, we can observe relationships between these variables and identify some differences. According to the plot, For classical, harmony *I-V-VI* has the highest distribution. In general, classical scores have bigger distributions for all 4 harmony levels. Also, harmony *I-V-VI*'s median seems to be higher around 7 for classical scores, whereas harmony *I-V-VI*'s median is around 5 for popular scores. In terms of instrument,string has the highest distribution, followed by piano and guitar, and guitar has the widest interval for classical music. Also, for each instrument, its third quartile seems to be at the median of the next highest instrument. For example, guitar's third quartile is approximately near piano's median, etc. For popular scores, everything is the opposite. String has the lowest distribution, followed by piano and guitar, and string has the widest interval. Also, for each instrument, its median seems to be at the third quartile of the next highest instrument. For example, guitar's median is approximately near piano's third quartile, etc. Finally for voice, all three levels seem to be evenly distributed for classical scores. However, for popular, the baseline level of contrary motion has the widest interval.

3.2 Examining the Effect of Experimental Factors on Classical Ratings

3.2.1 Interaction Only Model

We examine the influence of the three main experimental factors on classical music scores by creating a conventional linear model with all possible combination of interactions between the different instrument, harmony and voice levels. Moving on to variable selection, based on backwards selection AIC of the model for predicting classical music scores using instrument, harmony and voice, we see the selected variables are all the instrument, harmony and voice levels and both piano, strings, harmony *I-V-VI* are significant. In terms of interactions, we see that there are 6 interaction levels, all based on the interaction of harmony and voice. Of the different interaction levels, the interaction between Harmony *I-V-VI* and parallel 3rds seems to be most significant $\alpha = 0.05$ level. Thus, based on a rough examination of the effect of these factors with interactions, it seems instruments are quite significant, as well as the harmony level *I-V-VI*. Summary statistics of significant predictors in this model are shown in Figure 2 and section 2a (page 4) of the technical appendix.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.248	0.171	24.806	0.000
Instrumentpiano	1.374	0.113	12.184	0.000
Instrumentstring	3.133	0.112	27.947	0.000
HarmonyI-V-VI	1.140	0.225	5.074	0.000
HarmonyI-V-VI:Voicepar3rd	-0.684	0.318	-2.150	0.032

Figure 2: Coefficients of Significant Interactions in the Interaction-only Model

3.2.2 Adding Random Effects

Furthermore, we conduct an analysis of covariance test on the random effect of subjects, to examine whether random effects should be added to our model. Based on the results of the ANOVA test, we determine that we need the random effect since the coefficient of the random effect is statistically significant with a high test statistic of 763.59 and p value of 0. The mixed model with the random effect of subjected added portray similar results to the summary statistics in Figure 2. In terms of harmony, it seems harmony *I-V-VI* is significant and in terms of instruments, both piano and string instruments continue to be significant. The fixed effect of Voicepar3rd and Voicepar5th are also significant with high test statistics. These results seem to correspond to the background of the data in that the researchers' main hypotheses focused on Harmony *I-V-VI*, which is frequently rated as classical. The analysis and results are shown in Section 2b (pages 4-5) of the technical appendix.

Because the random intercept in a repeated measures model can account for "personal biases" in ratings, we decide it may be necessary to add more random effects of instrument, harmony and type of voice leading. For example, one participant may be more inclined to call music played by piano as more classical, and another person may be inclined to a specific type of harmony level. Thus, to account for these biases, we determine whether the model we fit above can be improved by adding one more of these random effects. To do this, we create multiple models with various combinations of random effects. Some models only have the random effect of instrument on subjects, other models have multiple random effects of instrument on subject and voice on subject, and one model includes all three random effects on subjects. By comparing each of these 7 models with the base model that only considers random effects on subjects, we determine that a mixed model with the fixed effects of instrument, harmony and voice, along with the addition of the random effects of instruments and harmony on subjects was most appropriate moving forward. Compared to other models, this model had the lowest AIC of 9953. Interestingly, the model that includes all three random effects on subjects had a very similar AIC. However, since the researchers hypothesize that the baseline voice leading level of contrary motion has the strong association with classical ratings, we determine it was not completely necessary to move forward with the random of effect of voice on subjects. Also, we are more interested in the effect of instrument and harmony, since we do not believe that the baseline level of these factors is most significant to classical ratings. The analysis and results are shown in Section 2c (pages 5-8) of the technical appendix.

3.2.3 Adding Additional Fixed Effects

After deciding our random effects, we then further proceed to determine which individual predictors should be added to the model as fixed effects. We add all variables to a full model and conduct backwards selection AIC to see which variables should be added as fixed effects. At first, 16 variables were selected when using conventional stepAIC on R. After the variables are selected, we add them as fixed effects to our mixed model for another variable selection method specialized for mixed models. Interestingly, no additional fixed effects were selected. Subsequently, there was no need to go back and check to see whether there should be a change in random effects. Since no new fixed effects were added, this finalized our model selection process. Additionally, there was no need to determine whether it seems to matter if the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits, since these variables were not selected in the variable selection process. We examine and interpret the coefficients, standard errors and test statistics for the baseline levels of each experimental factor. Summary statistics and the equation form of our final model are shown in Figures 3 and 4. The analysis and results are shown in Section 3 (pages 8-13) of the technical appendix.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Classical ~ Harmony + Instrument + Voice +
 (Harmony | Subject) + (Instrument | Subject)

     AIC      BIC   logLik deviance df.resid
  9959.3  10104.8  -4954.6   9909.3     2468

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.6963 -0.5796  0.0203  0.5478  5.9563

Random effects:
 Groups    Name            Variance Std.Dev. Corr
 Subject   (Intercept)     1.343042 1.1589
           HarmonyI-V-IV   0.030176 0.1737    0.61
           HarmonyI-V-VI   1.581470 1.2576   -0.46  0.23
           HarmonyIV-I-V   0.004251 0.0652    0.49  0.16  0.17
 Subject.1 (Intercept)     1.667561 1.2913
           Instrumentpiano 1.635053 1.2787   -0.59
           Instrumentstring 3.506574 1.8726  -0.87  0.66
 Residual                  2.456615 1.5674
Number of obs: 2493, groups:  Subject, 70

Fixed effects:
                Estimate Std. Error t value
(Intercept)      4.34349    0.22556  19.257
HarmonyI-V-IV   -0.03066    0.09119  -0.336
HarmonyI-V-VI    0.77007    0.17461   4.410
HarmonyIV-I-V    0.05199    0.08905   0.584
Instrumentpiano  1.36876    0.17141   7.985
Instrumentstring 3.12763    0.23661  13.218
Voicepar3rd     -0.40909    0.07694  -5.317
Voicepar5th     -0.36997    0.07690  -4.811

Correlation of Fixed Effects:
            (Intr) HI-V-I HI-V-V HIV-I- Instrmntp Instrmnts Vcpr3r
HrmnyI-V-IV -0.105
HrmnyI-V-VI -0.345  0.292
HrmnyIV-I-V -0.169  0.487  0.265
Instrumntpn -0.434  0.000  0.000 -0.001
Instrmntstr -0.618  0.000  0.000  0.000  0.631
Voicepar3rd -0.170 -0.002  0.000  0.002 -0.001     0.000
Voicepar5th -0.170 -0.002 -0.002 -0.001 -0.001     0.000     0.500
convergence code: 0
Model failed to converge with max|grad| = 0.116439 (tol = 0.002, component 1)
```

Figure 3: Summary Statistics of Final Model for Classical Ratings

$$\text{Classical} = \alpha_{0j[i]} + \alpha_{kj[i]}^{\text{Instrument}} + \alpha_{kj[i]}^{\text{Harmony}} + \alpha_{kj[i]}^{\text{Voice}} + \epsilon_i$$

$$\alpha_{0j[i]} = 4.34 + \eta_{0j} \sim (0, 1.26)$$

$$\alpha_{\text{Instrument: Piano}} = 1.37 + \eta_{1j} \sim (0, 1.63)$$
$$\alpha_{\text{Instrument: String}} = 3.12 + \eta_{2j} \sim (0, 3.50)$$
$$\alpha_{\text{Harmony I-V-IV}} = -0.031 + \eta_{2j} \sim (0, 0.03)$$

$$\alpha_{\text{Harmony I-V-VI}} = 0.77 + \eta_{2j} \sim (0, 1.58)$$

$$\alpha_{\text{Harmony IV-I-V}} = 0.052 + \eta_{2j} \sim (0, 0.0034)$$

$$\alpha_{\text{Voice Voice3rd}} = -0.41$$
$$\alpha_{\text{Voice Voice5th}} = -0.37$$

Figure 4: Mathematical Form of Final Model for Classical Ratings

- Interpretation of Coefficients Excluding Baseline Levels
  - Holding all variables constant, classical rating scores were on average 1.37 points higher for piano than electric guitars.
  - Holding all variables constant, classical rating scores were on average 3.12 points higher for strings than electric guitars.
  - Holding all variables constant, classical rating scores were on average 0.031 points lower for Harmony *I-V-IV* than *I-VI-V*.
  - Holding all variables constant, classical rating scores were on average 0.053 points higher for Harmony *IV-I-V* than *I-VI-V*.
  - Holding all variables constant, classical rating scores were on average 0.77 points higher for Harmony *I-V-IV* than *I-VI-V*.
  - Holding all variables constant, classical rating scores were on average 0.41 points lower for Voicepar 3rd than contrary motion.
  - Holding all variables constant, classical rating scores were on average 0.37 points lower for Voicepar 5th than contrary motion.

- Interpretation of Coefficients for Baseline Levels
  - Holding all variables constant, classical rating scores were on average 4.34 high for the baseline level electrical guitars.
  - Holding all variables constant, classical rating scores were on average 4.34 high for the baseline harmonic motion level *I-VI-V*.
  - Holding all variables constant, classical rating scores were on average 6.18 high for the baseline voice level contrary motion.

In terms of variance of random effects, piano had a variance of 1.63 and strings 3.52. For all harmony levels, the variances were lower than 1 and the residual variance of the random effects was 2.46. It seems that instruments has a much higher variance than harmony levels, and strings has approximately twice the variance of piano.

In terms of significant variables, *I-V-VI*, piano, string, Voicepar3rd and Voicepar5th seemed to be all significant with high test statistocs according to the summary statistics. We further examined all levels (including baseline) for the three factors while having incorporated random effects. Summary statistics of the coefficients of all the experimental fixed effects while considering random effects are shown in Figure 5. Based on the results, we conclude that all three instrument levels are strongly associated with high test statistics and low p-values, as seen in Figures 3, 5 and section 3c of the technical appendix. For harmonic motion, *I-V-VI* has the highest test statistic of 0.77 in Figure 3, whereas other motions had test statistics lower than 0.1. When looking at just fixed effects having incorporated random effects, *I-V-VI* still has the highest test statistic of 21.97, making it most significant. Finally, after having separately examined the effect of the voice levels, contrary motion does seem to have the strongest influence, as it has the highest test statistic among the three of approximately 6.18 in Figure 5.

	Estimate	Std. Error	t value
Instrumentguitar	4.343	0.210	20.710
Instrumentpiano	5.713	0.213	26.856
Instrumentstring	7.471	0.214	34.977

	Estimate	Std. Error	t value
HarmonyI-IV-V	4.343	0.225	19.271
HarmonyI-V-IV	4.313	0.234	18.423
HarmonyI-V-VI	5.114	0.233	21.965
HarmonyIV-I-V	4.395	0.228	19.285

	Estimate	Std. Error	t value
Voicecontrary	6.183	0.177	34.971
Voicepar3rd	5.775	0.177	32.659
Voicepar5th	5.814	0.177	32.869

Figure 5: Coefficients of All Experimental Factor Levels in Final Model

Conditional residuals show that the residuals are relatively well distributed, as it is difficult to see a subject with residuals deviating significantly. Conditional residuals of the final model are shown in Figure 6.

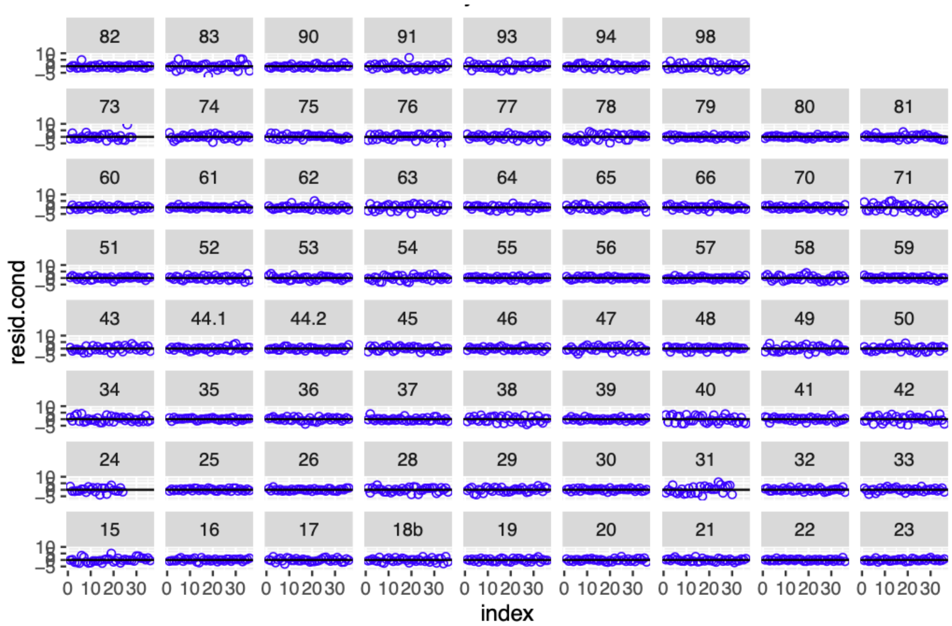


Figure 6: Conditional Residuals Across Subjects in Final Model

Thus, the results showed that all instruments are indeed strongly associated with classical ratings. Among the 4 harmony levels, *I-V-VI* has the strongest association with classical ratings. Also, both voice levels of Parallel 3rds and 5ths seem to strongly associated with classical ratings, but not as strong as contrary motion. Interestingly, when selecting random effects in our model, the random effect of voice on subjects was not significant enough to include in our final model.

3.3 Differences in Musicians and Non-Musicians Identifying Classical Music

We create 3 different thresholds for determining musicians and non-musicians. By dichotomizing the 'selfdeclare' variable into 3 ways, we can see which method is most effective and how sensitive our results can change. The first method was to set a score of 1 as a non-musician and 2 to 6 as musicians. The second method was to set a score of 1 to 3 as a non-musician and 4 to 6 as musicians. The third method was to set a score of 1 to 5 as a non-musician and 6 as musicians. We then add different dichotomized self-declared variables as interactions with each of the three experimental factors to the model selected in section 3.2.

Furthermore, we conduct ANOVA tests with our final model in section 3.2 as our null model to see which dichotomization is most significant. Based on dichotomization of different methods, it seems dividing the scores 1-3 to not declared and 4-6 to declared seems to have the most significant interactions based on the ANOVA tests. We included interactions for the dichotomized variables on self declare and instrument, harmony and subject.

To further see which interaction is most significant, we conduct more ANOVA tests, comparing the final model chosen in part 3.2 and models with interactions. Among all the interactions between the dichotomized variable for self declare, the interaction between the dichotomized variable and the harmony is most significant, with a high test statistic of 14.27 and a low p-value of 0.0067. The other interactions do not seem to be significant as they have low test statistics and high p-values according to the results of the ANOVA test. Thus, the results are indeed sensitive to how we divide 'selfdeclare.' Finally, it seems that there is a difference in how musicians and non-musicians identify classical music, especially with regard to harmony, as this interaction was most significant. The analysis is shown in the technical appendix section 4 (pages 13-17).

3.4 Differences in the Effect of Experimental Factors on Popular Ratings

In this section, we duplicate the methods in section 3.2 and 3.3, only with popular ratings as the response variable. Based on backwards selection AIC of the model for predicting popular music scores using instrument, harmony and voice, we see the significant variables are only the instrument, and harmony levels, not voice. In terms of interactions, we see that there are no significant interactions. Summary statistics of the interaction only model for popular ratings is shown in Figure 7.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.693	0.111	60.541	0.000
Instrumentpiano	-0.952	0.111	-8.572	0.000
Instrumentstring	-2.612	0.110	-23.662	0.000
HarmonyI-V-VI	-0.268	0.128	-2.097	0.036

Figure 7: Coefficients of Significant Interactions in the Interaction-only Model

Based on the results of the ANOVA test, we see that the coefficient of the random effect is statistically significant. For popular ratings, there was no need to create many combinations of random effects on subjects, as voice was not a significant variable. Based on the results of the ANOVA test, it seems necessary to choose the model with the random effect of instrument and harmony on subjects as it has the lowest AIC of 9999. Thus, it seems necessary to include the random effects of instrument and harmony on the subject. The analysis is shown in the technical appendix section 5a (pages 17-20).

Similarly to section 3.2, we decide it may be necessary to add more random effects of instrument and harmony. We conduct variable selection again, and according to the results of AIC variable selection, all additional fixed effects added were not selected, meaning that the only remaining variables were instrument. Neither harmony nor voice was selected. Thus, there was no need to add the random effect of harmony on subjects anymore, as the fixed effect of harmony is no longer influenced in the final model chosen by AIC. Thus, according to our final model for popular ratings, harmony and voice do not seem to be associated. For random effects, we only include the random effect of instruments on subjects. Summary statistics of our final model are shown in Figure 8.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Popular ~ Instrument + (Instrument | Subject)
   Data: df

     AIC      BIC   logLik deviance df.resid
 10104.1  10162.3  -5042.0  10084.1     2483

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8942 -0.6005 -0.0052  0.5851  5.7291

Random effects:
 Groups   Name             Variance Std.Dev. Corr
 Subject  (Intercept)      1.544    1.243
          Instrumentpiano  1.368    1.170    -0.25
          Instrumentstring 3.295    1.815    -0.40  0.73
 Residual                  2.826    1.681
Number of obs: 2493, groups:  Subject, 70

Fixed effects:
                 Estimate Std. Error t value
(Intercept)        6.5679     0.1596  41.152
Instrumentpiano   -0.9471     0.1626  -5.823
Instrumentstring  -2.6059     0.2320 -11.231

Correlation of Fixed Effects:
            (Intr) Instrmntp
Instrumntpn -0.335
Instrmntstr -0.440  0.674
convergence code: 0
Model failed to converge with max|grad| = 0.0178133 (tol = 0.002, component 1)
```

Figure 8: Summary Statistics of Final Model for Popular Ratings

$$\text{Popular} = \alpha_{0j[i]} + \alpha_{kj[i]}^{\text{Instrument}} + \epsilon_i$$

$$\alpha_{0j[i]} = 6.56 + \eta_{0j} \sim (0, 1.55)$$

$$\alpha_{\text{Instrument: Piano}} = -0.95 + \eta_{1j} \sim (0, 1.37)$$
$$\alpha_{\text{Instrument: String}} = -2.61 + \eta_{2j} \sim (0, 3.30)$$

Figure 9: Mathematical Form of Final Model for Popular Ratings

- Interpretation of Coefficients Excluding Baseline Levels
  - Holding all variables constant, popular rating scores were on average 0.95 points lower for piano than electric guitars.
  - Holding all variables constant, popular rating scores were on average 2.61 points lower for strings than electric guitars.
- Interpretation of Coefficients for Baseline Levels
  - Holding all variables constant, popular rating scores were on average 6.56 high for the baseline electrical guitars.

In terms of significant variables, piano and string seemed are both significant with high test statistics as shown in Figure 8. However, these results are different from predicting classical scores, as harmony and voice were also considered significant before. We further examined all levels (including baseline) for instrument while having incorporated random effects. Summary statistics of the coefficients of all the experimental fixed effects while considering random effects are shown in Figure 10. Similarly to classical ratings, we conclude that all three instrument levels are strongly associated with high test statistics as shown in Figure 10. The baseline guitar had the highest test statistic of 41.17, and these high test statistics show that all levels of instruments are strongly associated with popular ratings. In terms of variance of random effects, piano had a variance of 1.37 and strings 3.29. The residual variance of the random effects was 2.83. It seems that strings has approximately 2.5 times the variance of piano. The analysis for popular ratings is shown in the technical appendix section 5b (pages 20-24).

	Estimate	Std. Error	t value
Instrumentguitar	6.568	0.160	41.175
Instrumentpiano	5.621	0.186	30.261
Instrumentstring	3.962	0.216	18.312

Figure 10: Coefficients of All Experimental Factor Levels in Final Model for Popular Ratings

Conditional residuals show that the residuals are relatively well distributed, as it is difficult to see a subject with residuals deviating significantly. Conditional residuals of the final model are shown in Figure 11.

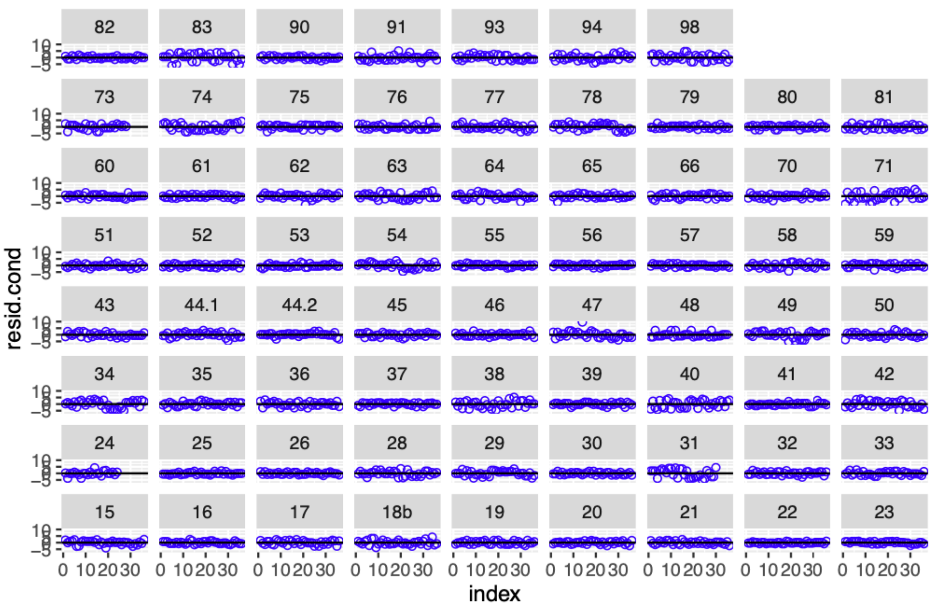


Figure 11: Conditional Residuals Across Subjects in Final Model for Popular Ratings

Similarly to the methods in section 3.3, we create 3 different thresholds for determining musicians and non-musicians. Based on dichotomization of different methods, it seems dividing the scores 1 to not declared and 2-6 to declared and 1-3 not declared and 4-6 declared seem to both have the most significant interactions based on the ANOVA tests. This was different from classical scores as only dichtomizing 1-3 not declared and 4-6 declared seemed to be significant according to the ANOVA test. We included interactions for the dichotomized variables on self declare and Instrument, harmony and subject. To further see which interaction is most significant, we conduct more ANOVA tests, comparing the final model chosen above and models with interactions. For consistency with the results in classical ratings, we will choose the second dichotomization method (1-3 not declared and 4-6 declared). Among all the interactions between the dichotomized variable for self declare, the interaction between the dichotomized variable and the harmony is most significant, similar to classical ratings. The other interactions do not seem to be significant according to the results of the ANOVA test. The analysis is shown in the technical appendix section 5c (pages 25-28).

4 Discussion

Having examined the data from the experiment, we conclude the following on experimental factors influence for classical and popular ratings. First, instrument does exert the strongest influence among the three design factors as it appears in both models for predicting classical and popular ratings. Among the levels of harmonic motion, *I-V-VI* has the strongest association with classical ratings, but not for popular ratings, as harmony does not have a significant association with popular ratings. Since our final model in section 3.2 did not include any of the Pachelbel rants/comedy bits, it seems that it does not matter whether the respondent is familiar with these bits to identify classical music. Finally, among the levels of voice leading, it seems both Parallel 3rds and 5ths are not as influential as contrary motion. However, voice leading is not an influential factor for popular ratings, as this variable was excluded from the model with popular ratings as the response. To summarize our findings, instruments are strongly associated with classical and popular ratings, as researchers suspected. The harmony level *I-V-VI* does have the strongest association with classical music, but not associated with popular. Finally, contrary motion has the strongest association with classical music, but voice levels are not associated with identifying popular music.

Having dichtomized the variable for one's declaration of musician status into three different methods, we determine that there are differences in the way that musicians and non-musicians identify classical music. We noted that these differences can also depend on the dichotomization threshold, as some bounds were more significant than others. Such was also the case for popular music.

We faced some limitations when analyzing the data. First, the data contains too many missing variables. When we initially removed all NA values in the data, more than 40% of the observations were removed. Thus, median imputation seemed to be an appropriate choice; however, this method can still skew the data and does not account for uncertainty in the imputations. However, this option was still preferable to removing all missing values in the data. Also, given that with the exception of the factor variables, all additional fixed effects in our final model were not chosen in our variable selection procedure, it seemed that many of the variables in the data were unnecessary and external to our analysis.

Future studies should consider using data without many missing values. Having a data set with approximately 40% of observations containing at least one missing value does not seem to be a practical start for analysis. Also, we believe that the study should incorporate different predictors for future experiments. Instead of asking many irrelevant survey questions that actually were insignificant in the modeling process, we can have variables that not only directly pertain to the three experimental factors, but also directly influence one's impact on instrument, harmony and voice leading. For example, the study can ask questions on whether participants are familiar or how familiar they are with music editing software such as GarageBand. Additionally, the study can divide the participants into different categories and conduct the experiment on separate pools. For example, one study can involve participants advanced music background, another can involve those with intermediate backgrounds, and etc. Nevertheless, the most important room for improvement in this study is to avoid utilizing a data set with more than 1000 missing values.

References

Sheather, S.J. (2009), *A Modern Approach to Regrssion with R*, New York: Springer Science + Business Media LLC

Technical Appendix

```
df <- read.csv('ratings.csv', header=TRUE)
library(plyr)
library(ggplot2)
library(arm)
library(lme4)
library(LMERConvenienceFunctions)
library(RLRsim)
library(sjPlot)
library(sjmisc)
library(sjlabelled)
library(kableExtra)

source('residual-functions.r')
```

Section 1 (Data Cleaning & Exploratory Data Analysis)

```
# Examine variables
#str(df)

# Data Cleaning

# Remove unneeded variables (first12 and X)

df$X = NULL
df$first12 = NULL

# Median imputation for observations with NA values
df$ConsNotes=ifelse(is.na(df$ConsNotes),median(df$ConsNotes,na.rm=T),df$ConsNotes)
df$PachListen=ifelse(is.na(df$PachListen),median(df$PachListen,na.rm=T),df$PachListen)
df$ClsListen=ifelse(is.na(df$ClsListen),median(df$ClsListen,na.rm=T),df$ClsListen)
df$KnowRob=ifelse(is.na(df$KnowRob),median(df$KnowRob,na.rm=T),df$KnowRob)
df$KnowAxis=ifelse(is.na(df$KnowAxis),median(df$KnowAxis,na.rm=T),df$KnowAxis)
df$X1990s2000s=ifelse(is.na(df$X1990s2000s),median(df$X1990s2000s,na.rm=T),df$X1990s2000s)
df$X1990s2000s.minus.1960s1970s=ifelse(is.na(df$X1990s2000s.minus.1960s1970s),
                                       median(df$X1990s2000s.minus.1960s1970s,na.rm=T),
                                       df$X1990s2000s.minus.1960s1970s)
df$CollegeMusic=ifelse(is.na(df$CollegeMusic),median(df$CollegeMusic,na.rm=T),df$CollegeMusic)
df$NoClass=ifelse(is.na(df$NoClass),median(df$NoClass,na.rm=T),df$NoClass)
df$APTheory=ifelse(is.na(df$ClsListen),median(df$ClsListen,na.rm=T),df$ClsListen)
df$Composing=ifelse(is.na(df$KnowRob),median(df$KnowRob,na.rm=T),df$KnowRob)

# Change numeric variables into factor variables
df$Selfdeclare = as.factor(df$Selfdeclare)

# Categorize ConsInstr
df$ConsInstr[which((df$ConsInstr == .67) | (df$ConsInstr == 1.33))] = 1
df$ConsInstr[which((df$ConsInstr == 1.67) | (df$ConsInstr == 2.33))] = 2
df$ConsInstr[which((df$ConsInstr == 2.67) | (df$ConsInstr == 3.33))] = 3
df$ConsInstr[which((df$ConsInstr == 3.67) | (df$ConsInstr == 4.33))] = 4
```

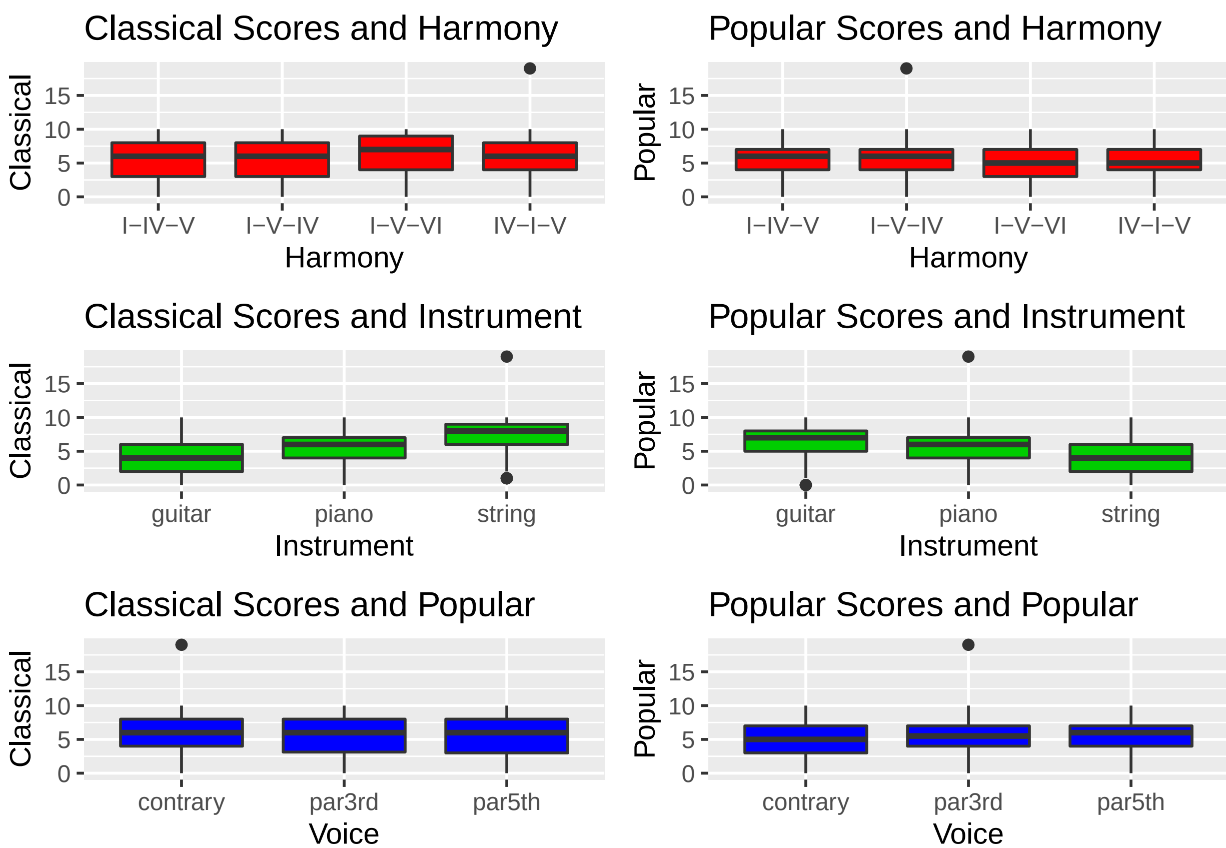
```r
df$ConsNotes[which(is.na(df$ConsNotes))] = 4
df$ConsInstr = as.factor(df$ConsInstr)
df$ConsNotes = as.factor(df$ConsNotes)
df$PachListen = as.factor(df$PachListen)
df$ClsListen = as.factor(df$ClsListen)
df$KnowRob = as.factor(df$KnowRob)
df$KnowAxis = as.factor(df$KnowAxis)
df$X1990s2000s = as.factor(df$X1990s2000s)
df$CollegeMusic = as.factor(df$CollegeMusic)
df$NoClass = as.factor(df$NoClass)
df$APTheory = as.factor(df$APTheory)
df$Composing = as.factor(df$Composing)
df$PianoPlay = as.factor(df$PianoPlay)
df$GuitarPlay = as.factor(df$GuitarPlay)

# Remove X1stInstr, X2ndInstr since there are too many NA values
df$X1stInstr = NULL
df$X2ndInstr = NULL

# Remove all NAs in Classical and Popular
df <- df[!is.na(df$Classical),]
df <- df[!is.na(df$Popular),]

# Check to see if there are any missing values remaining values
sum(is.na(df))
```

```
## [1] 0
```

```r
# Examine continous variables

par(mfrow=c(1,2))
#hist(df$OMSI)
#hist(log(df$OMSI))

# Log Transform OMSI since it is heavily skewed
df$OMSI = log(df$OMSI)
```

```r
library(gridExtra)

# Examine Boxplots of Harmony, instruments and voice on classical music

plot1 <- ggplot(df, aes(x = Harmony, y = Classical)) + geom_boxplot(fill=2) +
  labs(title = "Classical Scores and Harmony")

plot2 <- ggplot(df, aes(x = Instrument, y = Classical)) + geom_boxplot(fill=3) +
  labs(title = "Classical Scores and Instrument")

plot3 <- ggplot(df, aes(x = Voice, y = Classical)) + geom_boxplot(fill=4) +
  labs(title = "Classical Scores and Popular")


# Examine Boxplots of Harmony, instruments and voice on popular music

plot4 <- ggplot(df, aes(x = Harmony, y = Popular)) + geom_boxplot(fill=2) +
```

```r
  labs(title = "Popular Scores and Harmony")

plot5 <- ggplot(df, aes(x = Instrument, y = Popular)) + geom_boxplot(fill=3) +
  labs(title = "Popular Scores and Instrument")

plot6 <- ggplot(df, aes(x = Voice, y = Popular)) + geom_boxplot(fill=4) +
  labs(title = "Popular Scores and Popular")

grid.arrange(plot1, plot4, plot2, plot5, plot3, plot6, ncol=2)
```



- For classical, harmony I-V-VI has the highest distribution. In general, classical scores have bigger distributions for all 4 harmony levels. Also, harmony I-V-VI's median seems to be higher around 7 for classical scores, whereas harmony I-V-VI's median is around 5 for popular scores.
- For classical, string has the highest distribution, followed by piano and guitar, and guitar has the widest interval. Also, for each instrument, its 3rd quartile seems to be at the median of the next highest instrument. For example, guitar's 3rd quartile is approximately near piano's median, etc. For popular scores, everything is the opposite. String has the lowest distribution, followed by piano and guitar, and string has the widest interval. Also, for each instrument, its median seems to be at the 3rd quartile of the next highest instrument. For example, guitar's median is approximately near piano's 3rd quartile, etc.
- For voice, all three levels seem to be evenly distributed for classical scores. However, for popular, the baseline contrary has the widest interval.

Section 2 (Examining the Effect of the Three Experimental Factors on Classical Scores)

(a)

```
interaction_model <- lm(Classical ~ Instrument * Harmony * Voice, data = df)

int_mod_aic <- step(interaction_model, direction = "backward", trace = FALSE)

summary(int_mod_aic)$coefficients[c(1,2,3,5,10),]
```

```
##                             Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)                4.2484025  0.1712663 24.805826 1.597267e-121
## Instrumentpiano            1.3739814  0.1127683 12.184108  3.297153e-33
## Instrumentstring           3.1325594  0.1120878 27.947379 1.196103e-149
## HarmonyI-V-VI              1.1404386  0.2247809  5.073557  4.193656e-07
## HarmonyI-V-VI:Voicepar3rd -0.6839162  0.3180827 -2.150121  3.164210e-02
```

Based on backwards selection AIC of the model for predicting classical music scores using instrument, harmony and voice, we see the selected variables are all the instrument, harmony and voice levels and both instruments, harmony I-V-VI are significant. In terms of interactions, we see that there are 6 interaction levels, all based on the interaction of harmony and voice. Of the different interaction levels, the interaction between Harmony I-V-VI and parallel 3rds seems to be most significant at the 0.05 α level.

(b)

```
mod1 <- lm(Classical ~ Instrument + Harmony + Voice, data = df)
lmer.1 <- lmer(Classical ~ Instrument + Harmony + Voice + (1|Subject), data = df, REML=F)

anova(lmer.1, mod1)
```

```
## Data: df
## Models:
## mod1: Classical ~ Instrument + Harmony + Voice
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod1    9 11230 11283 -5606.2    11212
## lmer.1 10 10469 10527 -5224.4    10449 763.59      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the results of the ANOVA test, we see that the coefficient of the random effect is statistically significant with a high test statistic and a low p value. Thus, it seems we do need the random effect.

```
summary(lmer.1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##  10468.9  10527.1  -5224.4  10448.9     2483
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9844 -0.6404 -0.0182  0.6414  5.4886
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.677    1.295
##  Residual             3.571    1.890
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       4.34374    0.18809  23.094
## Instrumentpiano   1.37705    0.09305  14.800
## Instrumentstring  3.13161    0.09243  33.879
## HarmonyI-V-IV    -0.03251    0.10703  -0.304
## HarmonyI-V-VI     0.77096    0.10702   7.204
## HarmonyIV-I-V     0.04989    0.10694   0.467
## Voicepar3rd      -0.41506    0.09273  -4.476
## Voicepar5th      -0.37438    0.09268  -4.040
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.245
## Instrmntstr -0.246  0.498
## HrmnyI-V-IV -0.283  0.001    -0.001
## HrmnyI-V-VI -0.283  0.001    -0.001     0.499
## HrmnyIV-I-V -0.284 -0.001    -0.001     0.499  0.499
## Voicepar3rd -0.246 -0.001    -0.001    -0.002  0.001  0.002
## Voicepar5th -0.245 -0.001     0.000    -0.002 -0.003 -0.001  0.500
```

The results match with the summary statistics of the AIC selected model in part a, except now Voicepar3rd and Voicepar5th now seem to be significant with a very high test statistic. In terms of harmony, it seems harmony I-V-VI is significant and in terms of instruments, both piano and string instruments continue to be significant. These results seem to correspond to the background of the data in that the researchers' main hypotheses focused on Harmony I-V-VI, which is frequently rated as classical.

(c)

i.

I will conduct multiple ANOVA tests on each random effect on subject and select the model that has the lowest AIC and BIC.

```r
lmer.2 <- lmer(Classical ~ Instrument + Harmony + Voice + (Instrument|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.3 <- lmer(Classical ~ Instrument + Harmony + Voice + (Harmony|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.4 <- lmer(Classical ~ Instrument + Harmony + Voice + (Voice|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.5 <- lmer(Classical ~ Instrument + Harmony + Voice +
                 (Instrument|Subject) + (Harmony|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.6 <- lmer(Classical ~ Instrument + Harmony + Voice +
                 (Instrument|Subject) + (Voice|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)
```

```r
lmer.7 <- lmer(Classical ~ Instrument + Harmony + Voice +
                 (Harmony|Subject) + (Voice|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.8 <- lmer(Classical ~ Instrument + Harmony + Voice +
                 (Instrument|Subject) + (Harmony|Subject) + (Voice|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)
```

```r
anova(lmer.2, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.2: Classical ~ Instrument + Harmony + Voice + (Instrument | Subject)
##        Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10469 10527 -5224.4    10449
## lmer.2 15 10102 10190 -5036.1    10072 376.7      5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.3, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.3: Classical ~ Instrument + Harmony + Voice + (Harmony | Subject)
##        Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10469 10527 -5224.4    10449
## lmer.3 19 10388 10499 -5175.1    10350 98.67      9  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.4, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.4: Classical ~ Instrument + Harmony + Voice + (Voice | Subject)
##        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10469 10527 -5224.4    10449
## lmer.4 15 10479 10566 -5224.4    10449 0.0411      5          1
```

```r
anova(lmer.5, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.5: Classical ~ Instrument + Harmony + Voice + (Instrument | Subject) +
## lmer.5:     (Harmony | Subject)
##        Df     AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10468.9 10527 -5224.4  10448.9
## lmer.5 25  9959.3 10105 -4954.6   9909.3 539.59     15  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.6, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.6: Classical ~ Instrument + Harmony + Voice + (Instrument | Subject) +
## lmer.6:     (Voice | Subject)
##        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10469 10527 -5224.4    10449
## lmer.6 21 10114 10236 -5035.9    10072 377.13     11  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.7, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.7: Classical ~ Instrument + Harmony + Voice + (Harmony | Subject) +
## lmer.7:     (Voice | Subject)
##        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10469 10527 -5224.4    10449
## lmer.7 25 10400 10546 -5175.3    10350 98.332     15    2.7e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.8, lmer.1)
```

```
## Data: df
## Models:
## lmer.1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
## lmer.8: Classical ~ Instrument + Harmony + Voice + (Instrument | Subject) +
## lmer.8:     (Harmony | Subject) + (Voice | Subject)
##        Df     AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1 10 10468.9 10527 -5224.4  10448.9
## lmer.8 31  9969.5 10150 -4953.8   9907.5 541.36     21  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I choose 'lmer5' because it has the lowest AIC, BIC value of 9959.3 and 10105, respectively. Also, we saw from part a that voice is the least statistically significant variable among instrucment, harmony and voice. Thus, though the model for all three interactions has a low BIC and AIC, we saw that voice is not very useful to include in our model. Thus, according to our model, the best combinations of random effect terms are Instrument | Subject and Harmony | Subject.

ii.

```r
summary(lmer.5)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ Instrument + Harmony + Voice + (Instrument | Subject) +
##     (Harmony | Subject)
##    Data: df
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
```

```
##   9959.3  10104.8  -4954.6   9909.3     2468
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6943 -0.5799  0.0200  0.5473  5.9565
##
## Random effects:
##  Groups    Name             Variance Std.Dev. Corr
##  Subject   (Intercept)      1.262204 1.12348
##            Instrumentpiano  1.635332 1.27880  -0.67
##            Instrumentstring 3.505812 1.87238  -1.00  0.66
##  Subject.1 (Intercept)      1.744719 1.32088
##            HarmonyI-V-IV    0.029899 0.17291   0.55
##            HarmonyI-V-VI    1.578514 1.25639  -0.40  0.22
##            HarmonyIV-I-V    0.003423 0.05851   0.53 -0.01  0.17
##  Residual                   2.456564 1.56734
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       4.34349    0.22544  19.267
## Instrumentpiano   1.36874    0.17142   7.985
## Instrumentstring  3.12763    0.23659  13.220
## HarmonyI-V-IV    -0.03071    0.09116  -0.337
## HarmonyI-V-VI     0.77006    0.17449   4.413
## HarmonyIV-I-V     0.05198    0.08898   0.584
## Voicepar3rd      -0.40907    0.07694  -5.317
## Voicepar5th      -0.36997    0.07689  -4.811
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.434
## Instrmntstr -0.619  0.631
## HrmnyI-V-IV -0.103  0.000     0.000
## HrmnyI-V-VI -0.344  0.000     0.000     0.291
## HrmnyIV-I-V -0.167 -0.001     0.000     0.485  0.265
## Voicepar3rd -0.170 -0.001     0.000    -0.002  0.000  0.002
## Voicepar5th -0.170 -0.001     0.000    -0.002 -0.002 -0.001  0.500
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Section 3 (Adding Fixed Effects to the Model from Section 2)

(a)

```r
# Add other effects to the model from part b
# Use backward selection AIC to choose fixed effects

full.model <- lm(Classical ~ . -Subject - Popular, data = df)

full.model.aic <- stepAIC(full.model, direction = "backward", trace = FALSE)

sum.aic = summary(full.model.aic)
sum.aic$call
```

```
## lm(formula = Classical ~ Harmony + Instrument + Voice + Selfdeclare +
##     OMSI + ConsInstr + ConsNotes + Instr.minus.Notes + PachListen +
##     ClsListen + KnowAxis + X1990s2000s + X1990s2000s.minus.1960s1970s +
##     NoClass + PianoPlay + GuitarPlay, data = df)
```

```
# Final model

lmer.aic.model <- lmer(Classical ~ Harmony + Instrument + Voice + Selfdeclare +
                         OMSI + ConsInstr + ConsNotes + Instr.minus.Notes + PachListen +
                         ClsListen + KnowAxis + X1990s2000s + X1990s2000s.minus.1960s1970s +
                         NoClass + PianoPlay + GuitarPlay + (1|Subject),
                       data = df, REML = F, control = lmerControl(optimizer = 'bobyqa'))

# Final set of variables included in the model
sum.aic.lmer = summary(lmer.aic.model)

sum.aic.lmer$call
```

```
## lmer(formula = Classical ~ Harmony + Instrument + Voice + Selfdeclare +
##     OMSI + ConsInstr + ConsNotes + Instr.minus.Notes + PachListen +
##     ClsListen + KnowAxis + X1990s2000s + X1990s2000s.minus.1960s1970s +
##     NoClass + PianoPlay + GuitarPlay + (1 | Subject), data = df,
##     REML = F, control = lmerControl(optimizer = "bobyqa"))
```

(b)

```
sum.lmer = summary(lmer.fit.1)
sum.lmer$call
```

```
## lmer(formula = Classical ~ Harmony + Instrument + Voice + (1 |
##     Subject) + (Harmony | Subject) + (Instrument | Subject),
##     data = df, REML = TRUE, control = lmerControl(optimizer = "bobyqa"))
```

Based on the summary statistics, when we add the random effects of instrument, voice and harmony, the final model chosen by AIC adds the random effects of (Instrument|Subject) and (Harmony|Subject), but not (Voice|Subject). Additionally, according to the results of AIC variable selection, all fixed effects added in part a were not selected in this part, meaning that the only remaining variables are Harmony, Instrument and Voice. In choosing our final model, we will exclude the random effect of (1|Subject) since this is not needed. We are only interested in seeing the random effect of instrument and harmony on the subjects.

(c)

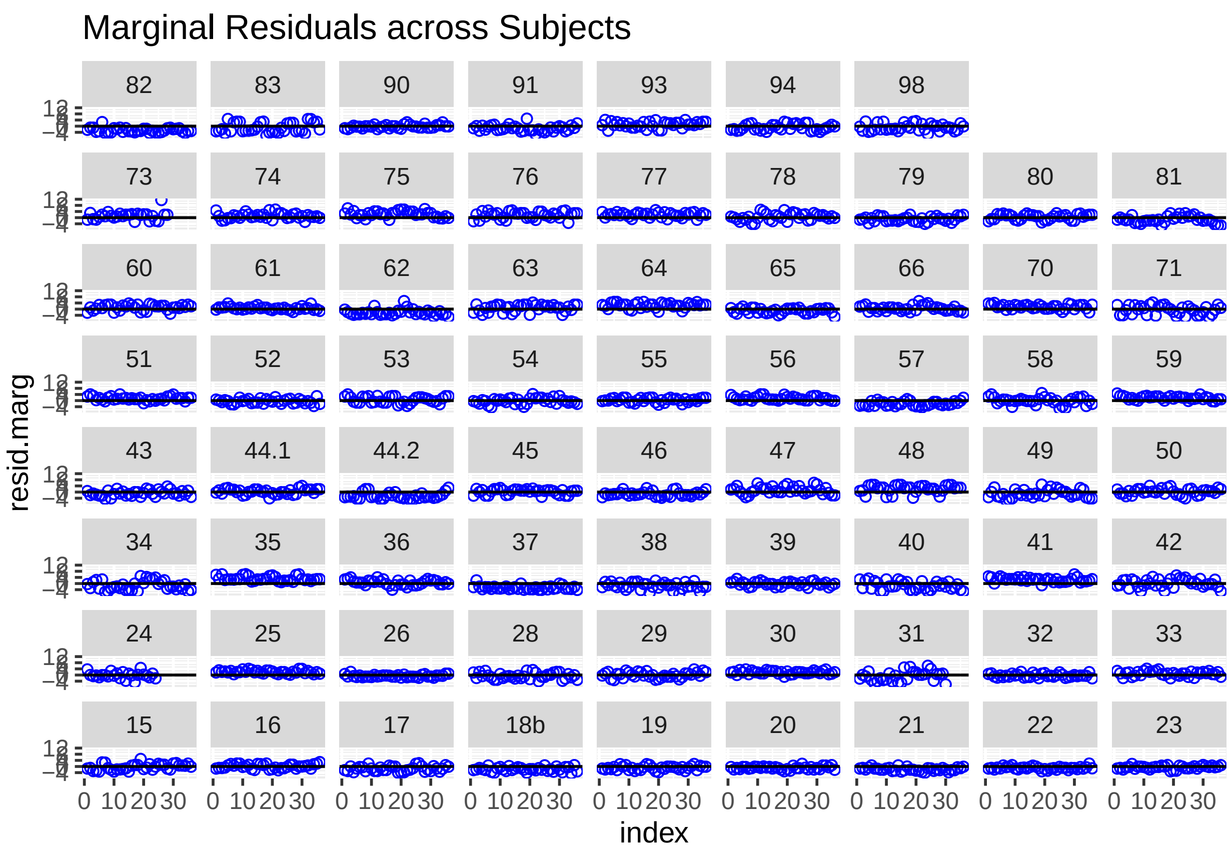
```
final.model <- lmer(Classical ~ Harmony + Instrument + Voice +
                      (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

summary(final.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
##     (Instrument | Subject)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##   9959.3  10104.8  -4954.6   9909.3     2468
##
## Scaled residuals:
```

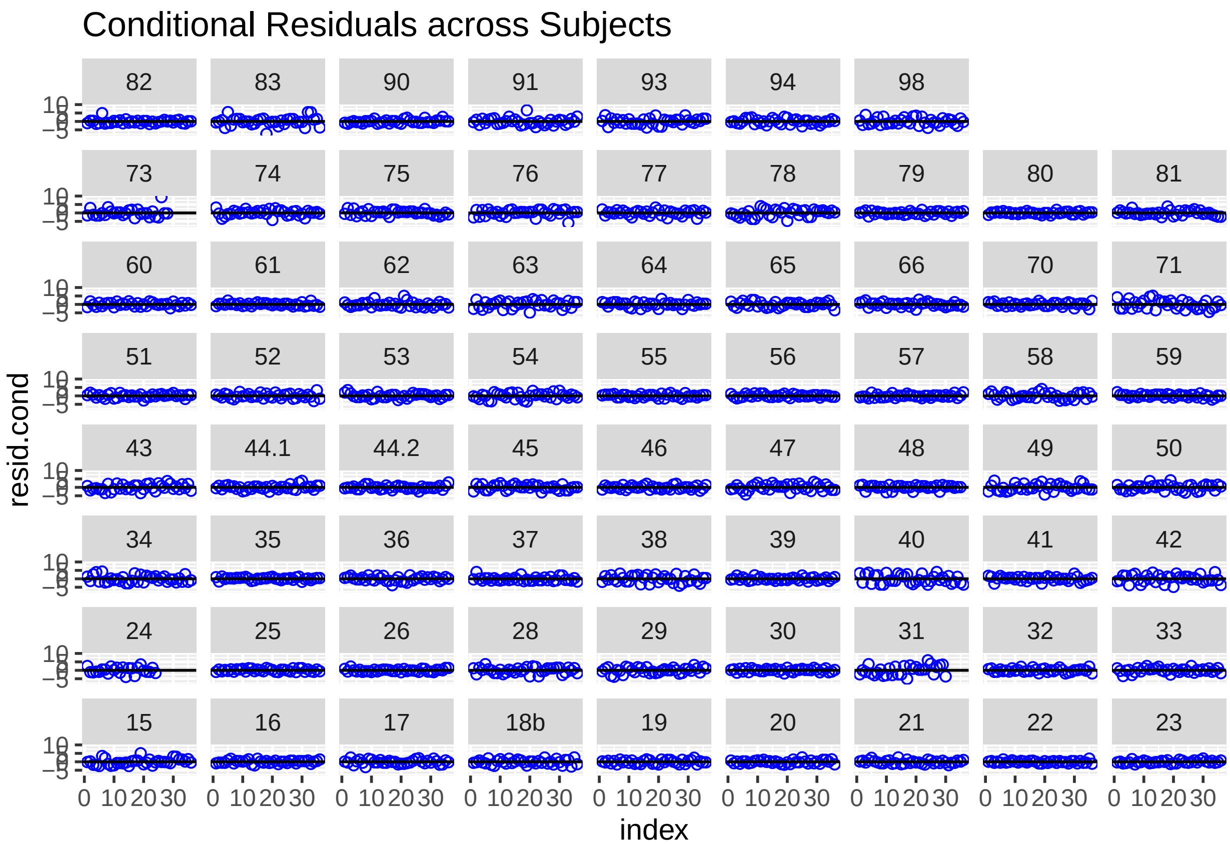
```
##      Min       1Q  Median      3Q     Max
## -4.6963 -0.5796  0.0203  0.5478  5.9563
##
## Random effects:
##  Groups    Name              Variance Std.Dev. Corr
##  Subject   (Intercept)       1.343042 1.1589
##            HarmonyI-V-IV     0.030176 0.1737    0.61
##            HarmonyI-V-VI     1.581470 1.2576   -0.46  0.23
##            HarmonyIV-I-V     0.004251 0.0652    0.49  0.16  0.17
##  Subject.1 (Intercept)       1.667561 1.2913
##            Instrumentpiano   1.635053 1.2787   -0.59
##            Instrumentstring  3.506574 1.8726   -0.87  0.66
##  Residual                    2.456615 1.5674
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       4.34349    0.22556  19.257
## HarmonyI-V-IV    -0.03066    0.09119  -0.336
## HarmonyI-V-VI     0.77007    0.17461   4.410
## HarmonyIV-I-V     0.05199    0.08905   0.584
## Instrumentpiano   1.36876    0.17141   7.985
## Instrumentstring  3.12763    0.23661  13.218
## Voicepar3rd      -0.40909    0.07694  -5.317
## Voicepar5th      -0.36997    0.07690  -4.811
##
## Correlation of Fixed Effects:
##             (Intr) HI-V-I HI-V-V HIV-I- Instrmntp Instrmnts Vcpr3r
## HrmnyI-V-IV -0.105
## HrmnyI-V-VI -0.345  0.292
## HrmnyIV-I-V -0.169  0.487  0.265
## Instrumntpn -0.434  0.000  0.000 -0.001
## Instrmntstr -0.618  0.000  0.000  0.000  0.631
## Voicepar3rd -0.170 -0.002  0.000  0.002 -0.001     0.000
## Voicepar5th -0.170 -0.002 -0.002 -0.001 -0.001     0.000     0.500
## convergence code: 0
## Model failed to converge with max|grad| = 0.116439 (tol = 0.002, component 1)
```

```r
# Residual Plots for Final Model

resid.marg <- r.marg(final.model)

resid.cond <- r.cond(final.model)

resid.reff <- r.reff(final.model)

subj <- as.numeric(df$Subject)
index <- subj
for (j in unique(subj)) {
  len <- sum(subj==j)
  index[subj==j] <- 1:len
}

# Marginal Residuals
```

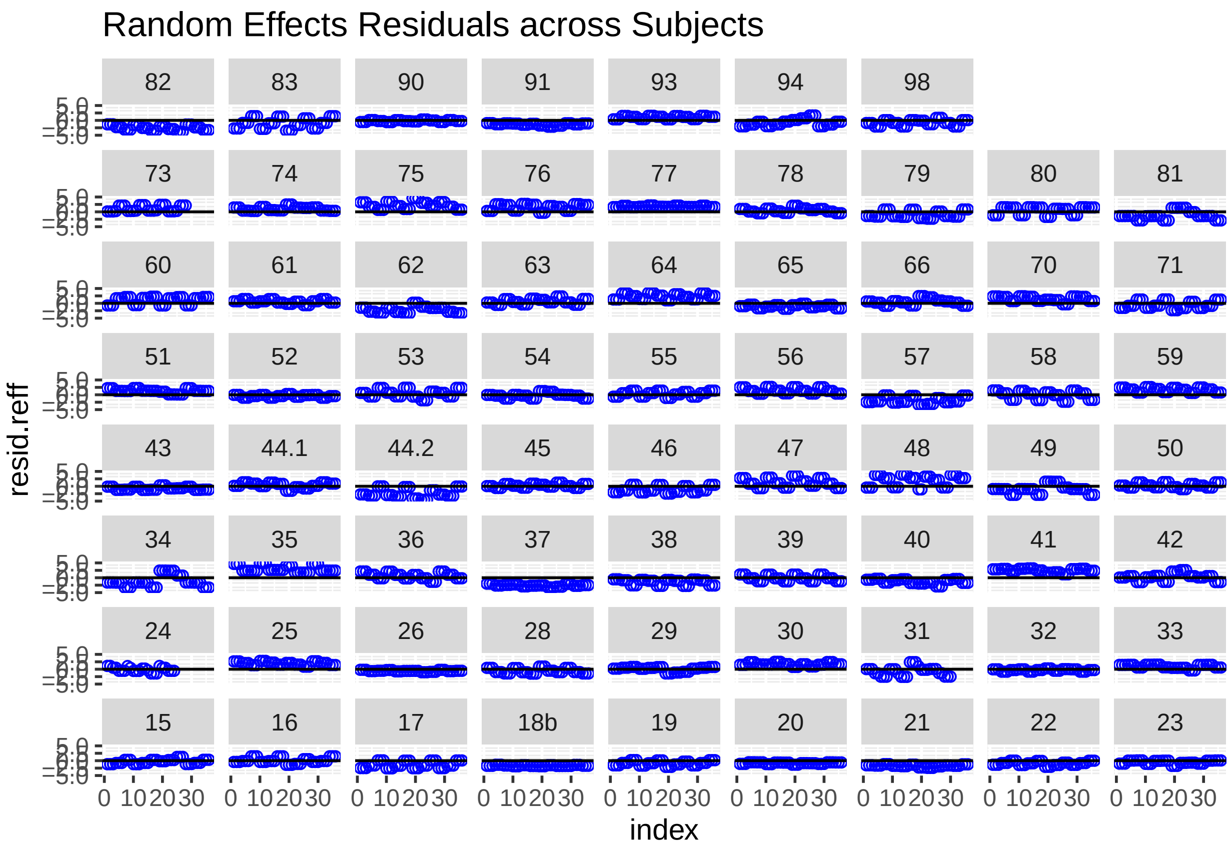
```r
new.data <- data.frame(index,resid.marg,df$Subject)
names(new.data) <- c("index","resid.marg","Subject")
ggplot(new.data,aes(x=index,y=resid.marg)) +
  facet_wrap( ~ Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
  geom_hline(yintercept=0) + labs(title = "Marginal Residuals across Subjects")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




```r
# Conditional Residuals

new.data <- data.frame(index,resid.cond,df$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) + labs(title = "Conditional Residuals across Subjects")
```




```
# Random Effects Residuals

new.data <- data.frame(index,resid.reff,df$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) + labs(title = "Random Effects Residuals across Subjects")
```



Based on the residual plots for marginal, conditional and random, it seems for most of the subjects, the residuals are rather distributed normally.

```
# Effect of fixed effects for all levels

final.model.eff.instr <- lmer(Classical ~ Instrument -1 + Harmony + Voice +
                          (Instrument - 1 + Harmony | Subject) +
                            (Instrument - 1 + Instrument | Subject), data = df, REML=F)

round(summary(final.model.eff.instr)$coefficients[c(1,2,3),],3)%>% kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = F)
```

	Estimate	Std. Error	t value
Instrumentguitar	4.343	0.210	20.710
Instrumentpiano	5.713	0.213	26.856
Instrumentstring	7.471	0.214	34.977

```
final.model.eff.harm <- lmer(Classical ~ Harmony -1 + Voice + Instrument +
                          (Harmony -1 + Harmony | Subject) + (Instrument | Subject), data = df, REML=F)
round(summary(final.model.eff.harm)$coefficients[c(1,2,3,4),],3)%>% kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = F)
```

	Estimate	Std. Error	t value
HarmonyI-IV-V	4.343	0.225	19.271
HarmonyI-V-IV	4.313	0.234	18.423
HarmonyI-V-VI	5.114	0.233	21.965
HarmonyIV-I-V	4.395	0.228	19.285

```
final.model.eff.voice <- lmer(Classical ~ Voice -1 + Harmony + Voice +
                      (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

round(summary(final.model.eff.voice)$coefficients[c(1,2,3),],3)%>% kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = F)
```

	Estimate	Std. Error	t value
Voicecontrary	6.183	0.177	34.971
Voicepar3rd	5.775	0.177	32.659
Voicepar5th	5.814	0.177	32.869

Section 4 (Dichotomizing self-declare for classical scores)

```
# Dichotomize self declare

# Dichotomization 1: if score 1 = not declared, 2 or above = declared

df$musician_status_1 = ifelse(as.numeric(df$Selfdeclare) <= 1 ,"0","1")


# Dichotomization 2: if score 1~3 = not declared, 4 or above = declared

df$musician_status_2 = ifelse(as.numeric(df$Selfdeclare) <= 3 ,"0","1")

# Dichotomization 3: if score 1~5 = not declared, 6 = declared

df$musician_status_3 = ifelse(as.numeric(df$Selfdeclare) <= 5 ,"0","1")


df$musician_status_1 = as.factor(df$musician_status_1)
df$musician_status_2 = as.factor(df$musician_status_2)
df$musician_status_3 = as.factor(df$musician_status_3)
```

```
# Add different dichotomized self-declared variables to the model from part c

# Dichotomization 1: if score 1 = not declared, 2 or above = declared

final.model.dichom.1 <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_1:Harmony +
                             musician_status_1:Instrument + musician_status_1:Voice +
                             (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)
```

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

```
# Conduct significance test with ANOVA to see which dichotomization is most significant

anova(final.model, final.model.dichom.1)
```

```
## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
## final.model.dichom.1: Classical ~ Harmony + Instrument + Voice + musician_status_1:Harmony +
## final.model.dichom.1:     musician_status_1:Instrument + musician_status_1:Voice +
## final.model.dichom.1:     (Harmony | Subject) + (Instrument | Subject)
##                       Df    AIC   BIC  logLik deviance  Chisq Chi Df
## final.model           25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.1 33 9960.8 10153 -4947.4   9894.8 14.488      8
##                       Pr(>Chisq)
## final.model
## final.model.dichom.1     0.0699 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Add different dichotomized self-declared variables to the model from part c

# Dichotomization 2: if score 1~3 = not declared, 4 or above = declared

final.model.dichom.2 <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_2:Harmony +
                               musician_status_2:Instrument + musician_status_2:Voice +
                               (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

# Conduct significance test with ANOVA to see which dichotomization is most significant

anova(final.model, final.model.dichom.2)
```

```
## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
## final.model.dichom.2: Classical ~ Harmony + Instrument + Voice + musician_status_2:Harmony +
## final.model.dichom.2:     musician_status_2:Instrument + musician_status_2:Voice +
## final.model.dichom.2:     (Harmony | Subject) + (Instrument | Subject)
##                       Df    AIC   BIC  logLik deviance Chisq Chi Df
## final.model           25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.2 33 9956.5 10149 -4945.3   9890.5 18.78      8
##                       Pr(>Chisq)
## final.model
## final.model.dichom.2    0.01608 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Add different dichotomized self-declared variables to the model from part c

# Dichotomization 3: if score 1~5 = not declared, 6 = declared

final.model.dichom.3 <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_3:Harmony +
                               musician_status_3:Instrument + musician_status_3:Voice +
                               (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

# Conduct significance test with ANOVA to see which dichotomization is most significant

anova(final.model, final.model.dichom.3)
```

```
## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
## final.model.dichom.3: Classical ~ Harmony + Instrument + Voice + musician_status_3:Harmony +
## final.model.dichom.3:     musician_status_3:Instrument + musician_status_3:Voice +
## final.model.dichom.3:     (Harmony | Subject) + (Instrument | Subject)
##                      Df    AIC   BIC  logLik deviance  Chisq Chi Df
## final.model          25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.3 33 9973.3 10165 -4953.7   9907.3 1.9467      8
##                      Pr(>Chisq)
## final.model
## final.model.dichom.3     0.9826
```

Based on dichotomization of different methods, it seems dividing the scores 1-3 to not declared and 4-6 to declared seems to have the most significant interactions based on the ANOVA tests. We included interactions for the dichotomized variables on self declare and Instrument, harmony and subject. To further see which interaction is most significant, we will conduct more ANOVA tests, comparing the final model chosen in part 3 and models with interactions.

```
final.model.dichom.2.harm <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_2:Harmony +
                               (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

## boundary (singular) fit: see ?isSingular

final.model.dichom.2.inst <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_2:Instrumen
                               musician_status_2:Instrument + musician_status_2:Voice +
                               (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

final.model.dichom.2.voice <- lmer(Classical ~ Harmony + Instrument + Voice + musician_status_2:Voice +
                               (Harmony | Subject) + (Instrument | Subject), data = df, REML=F)

## boundary (singular) fit: see ?isSingular

anova(final.model, final.model.dichom.2.harm)

## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
## final.model.dichom.2.harm: Classical ~ Harmony + Instrument + Voice + musician_status_2:Harmony +
## final.model.dichom.2.harm:     (Harmony | Subject) + (Instrument | Subject)
##                           Df    AIC   BIC  logLik deviance  Chisq Chi Df
## final.model               25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.2.harm 29 9953.1 10122 -4947.5   9895.1 14.207      4
##                           Pr(>Chisq)
## final.model
## final.model.dichom.2.harm   0.006662 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(final.model, final.model.dichom.2.inst)

## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
```

```
## final.model.dichom.2.inst: Classical ~ Harmony + Instrument + Voice + musician_status_2:Instrument +
## final.model.dichom.2.inst:     musician_status_2:Instrument + musician_status_2:Voice +
## final.model.dichom.2.inst:     (Harmony | Subject) + (Instrument | Subject)
##                             Df    AIC   BIC  logLik deviance  Chisq Chi Df
## final.model                 25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.2.inst 30 9965.7 10140 -4952.9   9905.7 3.5751      5
##                             Pr(>Chisq)
## final.model
## final.model.dichom.2.inst     0.6121
```

```
anova(final.model, final.model.dichom.2.voice)
```

```
## Data: df
## Models:
## final.model: Classical ~ Harmony + Instrument + Voice + (Harmony | Subject) +
## final.model:     (Instrument | Subject)
## final.model.dichom.2.voice: Classical ~ Harmony + Instrument + Voice + musician_status_2:Voice +
## final.model.dichom.2.voice:     (Harmony | Subject) + (Instrument | Subject)
##                              Df    AIC   BIC  logLik deviance  Chisq Chi Df
## final.model                  25 9959.3 10105 -4954.6   9909.3
## final.model.dichom.2.voice 28 9964.4 10127 -4954.2   9908.4 0.8417      3
##                              Pr(>Chisq)
## final.model
## final.model.dichom.2.voice     0.8395
```

It seems that among all the interactions between the dichotomized variable for self declare, the interaction between the dichotomized variable and the harmony is most significant, with a high test statistic and low p-value. The other interactions do not seem to be significant as they have low test statistics and high p-values according to the results of the ANOVA test.

Section 5 (Repeated Analysis for Popular Ratings)

(a)

```
# Observe the effect of interactions

interaction_model_pop <- lm(Popular ~ Instrument * Harmony * Voice, data = df)

int_mod_aic_pop <- step(interaction_model_pop, direction = "backward", trace = FALSE)

summary(int_mod_aic_pop)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony, data = df)
##
## 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
```

Based on backwards selection AIC of the model for predicting popular music scores using instrument, harmony and voice, we see the significant variables are only the instrument, and harmony levels, not voice. In terms of interactions, we see that there are no significant interactions.

- Test whether random intercept is needed in the model.

```r
mod1pop <- lm(Popular ~ Instrument + Harmony, data = df)
lmer.1.pop <- lmer(Popular ~ Instrument + Harmony + (1|Subject), data = df, REML=F)

anova(lmer.1.pop, mod1pop)
```

```
## Data: df
## Models:
## mod1pop: Popular ~ Instrument + Harmony
## lmer.1.pop: Popular ~ Instrument + Harmony + (1 | Subject)
##            Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod1pop     7 11142 11183 -5564.1    11128
## lmer.1.pop  8 10431 10477 -5207.4    10415 713.43      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the results of the ANOVA test, we see that the coefficient of the random effect is statistically significant. Thus, it seems we do need the random effect.

- Test to see which random effect is needed.

```r
lmer.2.pop <- lmer(Popular ~ Instrument + Harmony + (Instrument|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.3.pop <- lmer(Popular ~ Instrument + Harmony + (Harmony|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)

lmer.4.pop <- lmer(Popular ~ Instrument + Harmony +
                 (Instrument|Subject) + (Harmony|Subject), data = df,
               control = lmerControl(optimizer = 'bobyqa'), REML=F)
```

```r
anova(lmer.2.pop, lmer.1.pop)
```

```
## Data: df
## Models:
## lmer.1.pop: Popular ~ Instrument + Harmony + (1 | Subject)
## lmer.2.pop: Popular ~ Instrument + Harmony + (Instrument | Subject)
##            Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1.pop  8 10431 10477 -5207.4    10415
## lmer.2.pop 13 10099 10175 -5036.5    10073 341.66      5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.3.pop, lmer.1.pop)
```

```
## Data: df
## Models:
## lmer.1.pop: Popular ~ Instrument + Harmony + (1 | Subject)
## lmer.3.pop: Popular ~ Instrument + Harmony + (Harmony | Subject)
##            Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1.pop  8 10431 10477 -5207.4    10415
## lmer.3.pop 17 10384 10483 -5174.9    10350 64.899      9  1.509e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmer.4.pop, lmer.1.pop)
```

```
## Data: df
## Models:
## lmer.1.pop: Popular ~ Instrument + Harmony + (1 | Subject)
## lmer.4.pop: Popular ~ Instrument + Harmony + (Instrument | Subject) + (Harmony |
## lmer.4.pop:     Subject)
##            Df     AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer.1.pop  8 10430.7 10477 -5207.4  10414.7
## lmer.4.pop 23  9999.1 10133 -4976.6   9953.1 461.61     15  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the results of the ANOVA test, it seems necessary to choose lmer4 because it has the lowest AIC, BIC value. Thus, it seems necessary to include the random effects of instrument and harmony on the subject.

lmer.4: Popular ~ Instrument + Harmony + (Instrument | Subject) + (Harmony | Subject)

```r
summary(lmer.4.pop)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Popular ~ Instrument + Harmony + (Instrument | Subject) + (Harmony |
##     Subject)
##    Data: df
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   9999.1  10133.0  -4976.6   9953.1     2470
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9087 -0.5878  0.0143  0.5682  5.3717
##
## Random effects:
##  Groups    Name             Variance Std.Dev. Corr
##  Subject   (Intercept)      0.3225   0.5679
##            Instrumentpiano  1.4191   1.1913   -0.64
##            Instrumentstring 3.3504   1.8304   -0.91  0.72
##  Subject.1 (Intercept)      1.4059   1.1857
##            HarmonyI-V-IV    0.1113   0.3336    0.44
##            HarmonyI-V-VI    0.9014   0.9494   -0.29 -0.45
##            HarmonyIV-I-V    0.2500   0.5000   -0.34 -0.80 -0.18
##  Residual                   2.5073   1.5834
```

```
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       6.68764    0.17527  38.157
## Instrumentpiano  -0.94663    0.16252  -5.825
## Instrumentstring -2.60466    0.23211 -11.222
## HarmonyI-V-IV    -0.02553    0.09818  -0.260
## HarmonyI-V-VI    -0.27278    0.14468  -1.885
## HarmonyIV-I-V    -0.18550    0.10773  -1.722
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V
## Instrumntpn -0.322
## Instrmntstr -0.405  0.675
## HrmnyI-V-IV -0.089  0.000     0.000
## HrmnyI-V-VI -0.341  0.001     0.000     0.140
## HrmnyIV-I-V -0.365 -0.001     0.000     0.200  0.178
## convergence code: 1
## boundary (singular) fit: see ?isSingular
```

(b)

```r
# Add other effects to the model from part a
# Use backward selection AIC to choose fixed effects

full.model.pop <- lm(Popular ~ . -Subject - Classical, data = df)

full.model.aic.pop <- stepAIC(full.model.pop, direction = "backward", trace = FALSE)

sum.aic.pop = summary(full.model.aic.pop)
sum.aic.pop$call
```

```
## lm(formula = Popular ~ Harmony + Instrument + Selfdeclare + OMSI +
##     X16.minus.17 + ConsInstr + ConsNotes + Instr.minus.Notes +
##     PachListen + ClsListen + KnowRob + KnowAxis + X1990s2000s +
##     X1990s2000s.minus.1960s1970s + CollegeMusic + NoClass + PianoPlay +
##     GuitarPlay, data = df)
```

```r
# Final model

lmer.aic.model.pop <- lmer(Popular ~ Harmony + Instrument + Selfdeclare + OMSI +
                             X16.minus.17 + ConsInstr + ConsNotes + Instr.minus.Notes +
                             PachListen + ClsListen + KnowRob + KnowAxis + X1990s2000s +
                             X1990s2000s.minus.1960s1970s + CollegeMusic + NoClass + PianoPlay +
                             GuitarPlay + (1|Subject),
                       data = df, REML = F, control = lmerControl(optimizer = 'bobyqa'))

# Final set of variables included in the model

sum.aic.lmer.pop = summary(lmer.aic.model.pop)

sum.aic.lmer.pop$call
```

```
## lmer(formula = Popular ~ Harmony + Instrument + Selfdeclare +
##     OMSI + X16.minus.17 + ConsInstr + ConsNotes + Instr.minus.Notes +
```

```
##     PachListen + ClsListen + KnowRob + KnowAxis + X1990s2000s +
##     X1990s2000s.minus.1960s1970s + CollegeMusic + NoClass + PianoPlay +
##     GuitarPlay + (1 | Subject), data = df, REML = F, control = lmerControl(optimizer = "bobyqa"))
```

```r
sum.lmer.instr = summary(lmer.fit.instr)
sum.lmer.instr$call
```

```
## lmer(formula = Popular ~ Instrument + (1 | Subject) + (Instrument |
##     Subject), data = df, REML = TRUE, control = lmerControl(optimizer = "bobyqa"))
```

Based on the summary statistics, when we add the random effects of instrument, voice and harmony, the final model chosen by AIC adds the random effects of (Instrument|Subject), but not (Harmony|Subject) and (Voice|Subject). Additionally, according to the results of AIC variable selection, all fixed effects added in part a were not selected in this part, meaning that the only remaining variables are Instrument. When we did AIC variable selection in part a, we noticed that voice was not selected at all. Interestingly, according to our final model for predicting popular scores, harmony and voice are not needed. In choosing our final model, we will exclude the random effect of (1|Subject) since this is not needed. We are only interested in seeing the random effect of instrument on the subjects.

```r
final.model.pop <- lmer(Popular ~ Instrument + (Instrument | Subject), data = df, REML=F)

summary(final.model.pop)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ Instrument + (Instrument | Subject)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##  10104.1  10162.3  -5042.0  10084.1     2483
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8942 -0.6005 -0.0052  0.5851  5.7291
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      1.544    1.243
##           Instrumentpiano  1.368    1.170    -0.25
##           Instrumentstring 3.295    1.815    -0.40  0.73
##  Residual                  2.826    1.681
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)        6.5679     0.1596  41.152
## Instrumentpiano   -0.9471     0.1626  -5.823
## Instrumentstring  -2.6059     0.2320 -11.231
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp
## Instrumntpn -0.335
## Instrmntstr -0.440  0.674
## convergence code: 0
## Model failed to converge with max|grad| = 0.0178133 (tol = 0.002, component 1)
```

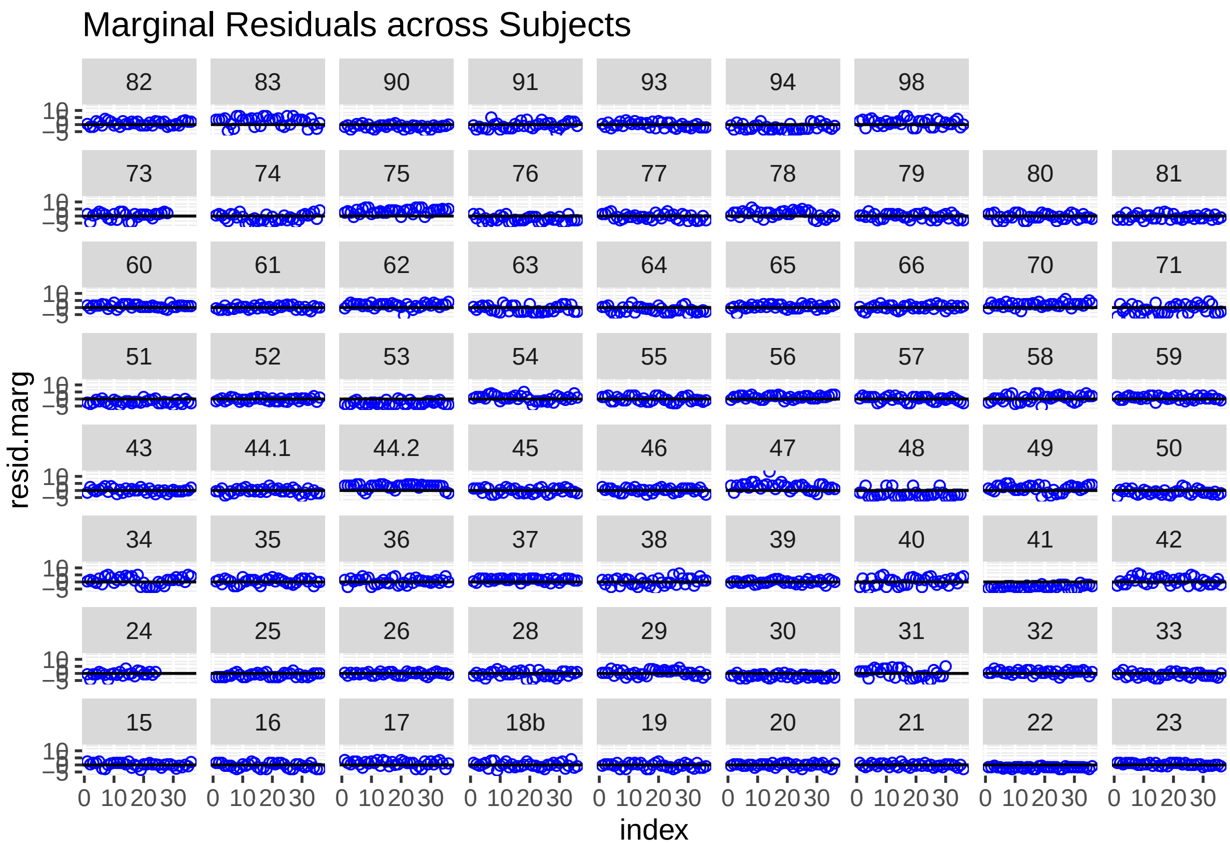
```r
final.mode.pop.eff.instr <- lmer(Popular ~ Instrument -1 +
                                   (Instrument -1 + Instrument | Subject), data = df, REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00639473
## (tol = 0.002, component 1)
```

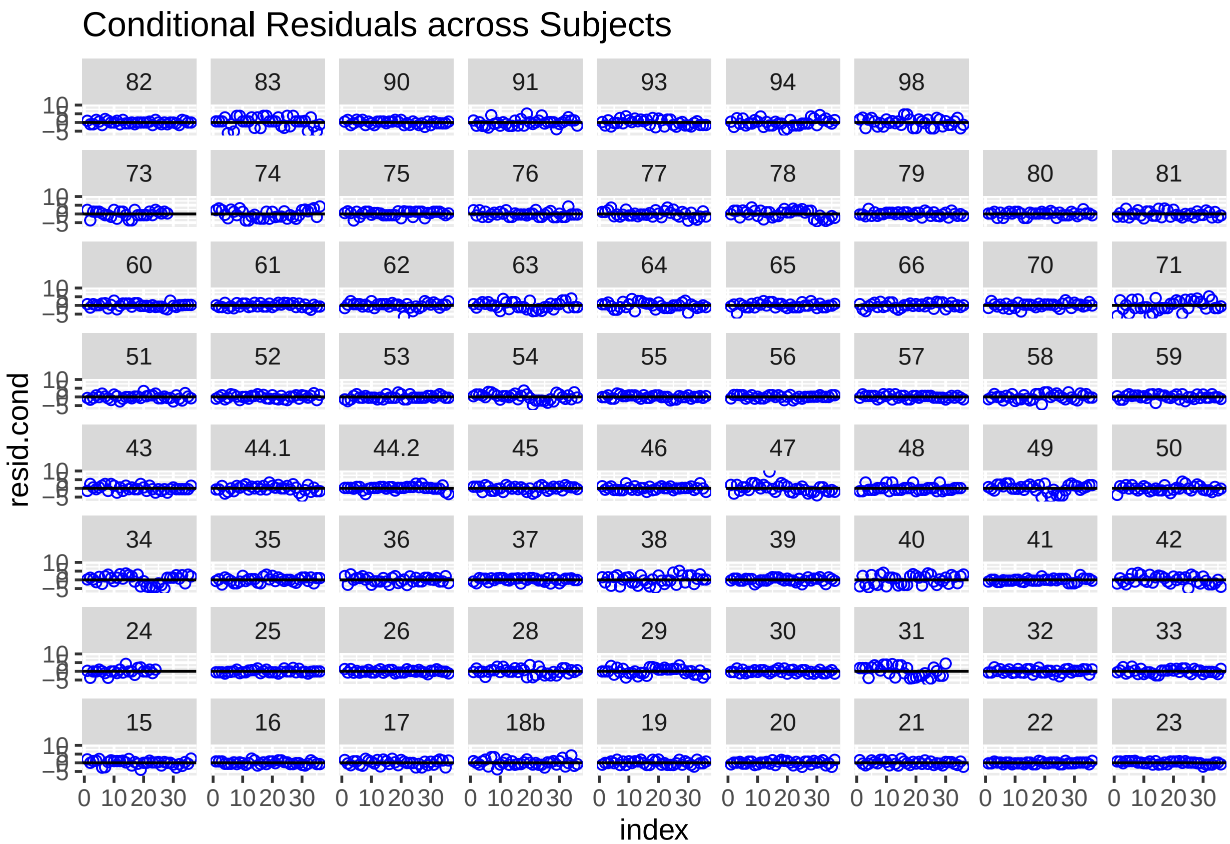
```r
round(summary(final.mode.pop.eff.instr)$coefficients[c(1,2,3),],3) %>% kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = F)
```

	Estimate	Std. Error	t value
Instrumentguitar	6.568	0.160	41.175
Instrumentpiano	5.621	0.186	30.261
Instrumentstring	3.962	0.216	18.312

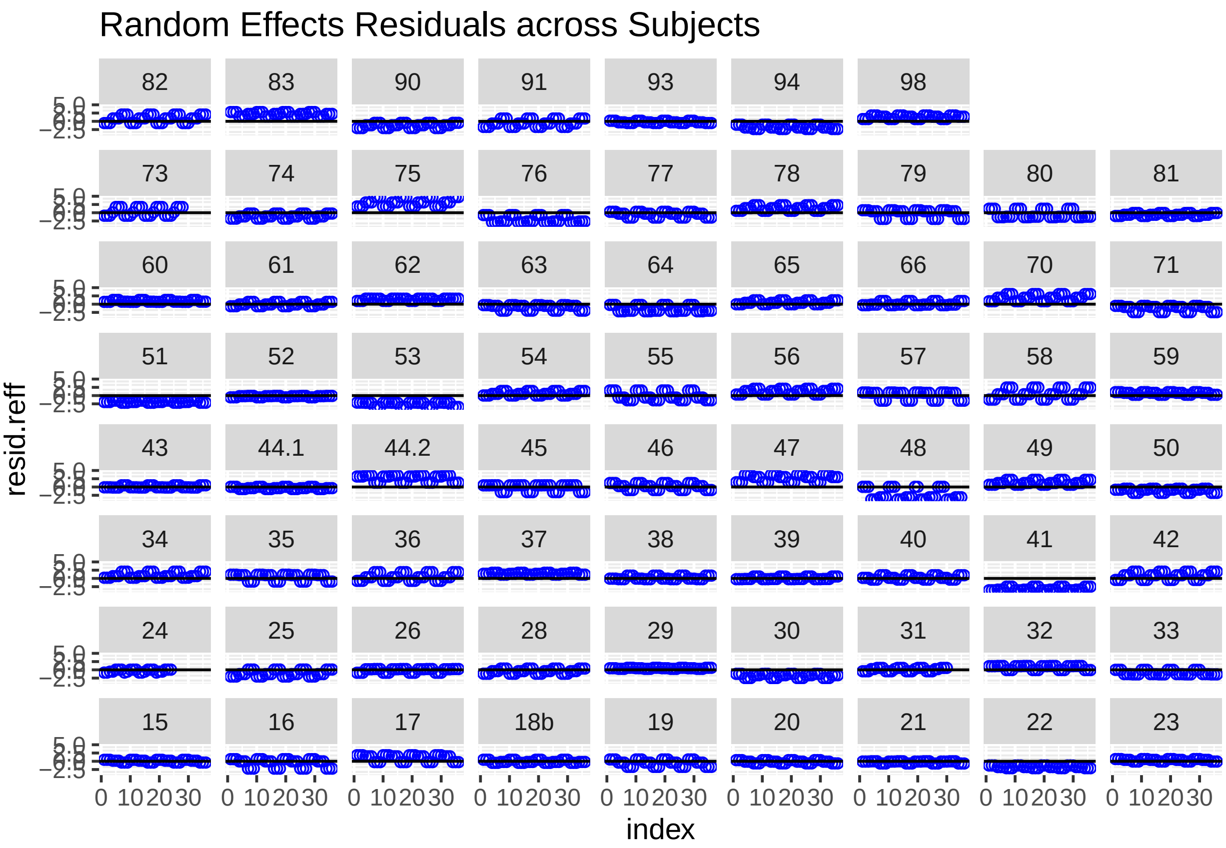
```r
# Residual Plots for Final Model

resid.marg <- r.marg(final.model.pop)

resid.cond <- r.cond(final.model.pop)

resid.reff <- r.reff(final.model.pop)

subj <- as.numeric(df$Subject)
index <- subj
for (j in unique(subj)) {
  len <- sum(subj==j)
  index[subj==j] <- 1:len
}

# Marginal Residuals

new.data <- data.frame(index,resid.marg,df$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) + labs(title = "Marginal Residuals across Subjects")
```




```r
# Conditional Residuals

new.data <- data.frame(index,resid.cond,df$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) + labs(title = "Conditional Residuals across Subjects")
```




```r
# Random Effects Residuals

new.data <- data.frame(index,resid.reff,df$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) + labs(title = "Random Effects Residuals across Subjects")
```



Based on the residual plots for marginal, conditional and random, it seems for most of the subjects, the residuals are rather distributed normally.

(c)

```
# Add different dichotomized self-declared variables to the model from part c

final.model.dichom.1.pop <- lmer(Popular ~ Instrument + musician_status_1:Harmony +
                              musician_status_1:Instrument + musician_status_1:Voice +
                              OMSI*musician_status_1 + (Instrument | Subject), data = df, REML=F)

anova(final.model.pop, final.model.dichom.1.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.1.pop: Popular ~ Instrument + musician_status_1:Harmony + musician_status_1:Instru
## final.model.dichom.1.pop:     musician_status_1:Voice + OMSI * musician_status_1 + (Instrument |
## final.model.dichom.1.pop:     Subject)
##                          Df   AIC   BIC  logLik deviance  Chisq Chi Df
## final.model.pop          10 10104 10162 -5042.0    10084
## final.model.dichom.1.pop 25 10076 10222 -5013.2    10026 57.734     15
##                          Pr(>Chisq)
## final.model.pop
## final.model.dichom.1.pop  6.158e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Add different dichotomized self-declared variables to the model from part c

final.model.dichom.2.pop <- lmer(Popular ~ Instrument + musician_status_2:Harmony +
                                 musician_status_2:Instrument + musician_status_2:Voice +
                                 OMSI*musician_status_2 + (Instrument | Subject), data = df, REML=F)

anova(final.model.pop, final.model.dichom.2.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.2.pop: Popular ~ Instrument + musician_status_2:Harmony + musician_status_2:Instru
## final.model.dichom.2.pop:     musician_status_2:Voice + OMSI * musician_status_2 + (Instrument |
## final.model.dichom.2.pop:     Subject)
##                          Df   AIC   BIC  logLik deviance  Chisq Chi Df
## final.model.pop          10 10104 10162 -5042.0    10084
## final.model.dichom.2.pop 25 10073 10219 -5011.6    10023 60.952     15
##                          Pr(>Chisq)
## final.model.pop
## final.model.dichom.2.pop  1.729e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Add different dichotomized self-declared variables to the model from part c

final.model.dichom.3.pop <- lmer(Popular ~ Instrument + musician_status_3:Harmony +
                                 musician_status_3:Instrument + musician_status_3:Voice +
                                 OMSI*musician_status_3 + (Instrument | Subject), data = df, REML=F)

anova(final.model.pop, final.model.dichom.3.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.3.pop: Popular ~ Instrument + musician_status_3:Harmony + musician_status_3:Instru
## final.model.dichom.3.pop:     musician_status_3:Voice + OMSI * musician_status_3 + (Instrument |
## final.model.dichom.3.pop:     Subject)
##                          Df   AIC   BIC logLik deviance  Chisq Chi Df
## final.model.pop          10 10104 10162  -5042    10084
## final.model.dichom.3.pop 24 10112 10252  -5032    10064 20.021     14
##                          Pr(>Chisq)
## final.model.pop
## final.model.dichom.3.pop     0.1295
```

Based on dichotomization of different methods, it seems dividing the scores 1 to not declared and 2-6 to declared and 1-3 not declared and 4-6 declared seem to both have the most significant interactions based on the ANOVA tests. This was different from classical scores as only dichtomizing 1-3 not declared and 4-6 declared seemed to be significant according to the ANOVA test. We included interactions for the dichotomized variables on self declare and Instrument, harmony and subject. To further see which interaction is most significant, we will conduct more ANOVA tests, comparing the final model chosen in part b and models with interactions. For consistency with the results in question 4, we will choose the second dichotomization method (1-3 not declared and 4-6 declared).

```r
# Test interactions on dichotomized selfdeclared and Harmony, Instrument, and Voice

final.model.dichom.2.harm.pop <- lmer(Popular ~ Instrument + musician_status_2:Harmony +
```

```r
                                    (Instrument | Subject), data = df, REML=F)

final.model.dichom.2.instr.pop <- lmer(Popular ~ Instrument + musician_status_2:Instrument +
                                    (Instrument | Subject), data = df, REML=F)

final.model.dichom.2.voice.pop <- lmer(Popular ~ Instrument + musician_status_2:Voice +
                                    (Instrument | Subject), data = df, REML=F)


anova(final.model.pop, final.model.dichom.2.harm.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.2.harm.pop: Popular ~ Instrument + musician_status_2:Harmony + (Instrument |
## final.model.dichom.2.harm.pop:     Subject)
##                               Df   AIC   BIC  logLik deviance  Chisq
## final.model.pop               10 10104 10162 -5042.0    10084
## final.model.dichom.2.harm.pop 17 10066 10166 -5016.3    10032 51.584
##                               Chi Df Pr(>Chisq)
## final.model.pop
## final.model.dichom.2.harm.pop      7  7.051e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(final.model.pop, final.model.dichom.2.instr.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.2.instr.pop: Popular ~ Instrument + musician_status_2:Instrument + (Instrument |
## final.model.dichom.2.instr.pop:     Subject)
##                                Df   AIC   BIC  logLik deviance  Chisq
## final.model.pop                10 10104 10162 -5042.0    10084
## final.model.dichom.2.instr.pop 13 10109 10185 -5041.5    10083 1.0102
##                                Chi Df Pr(>Chisq)
## final.model.pop
## final.model.dichom.2.instr.pop      3     0.7988
```

```r
anova(final.model.pop, final.model.dichom.2.voice.pop)
```

```
## Data: df
## Models:
## final.model.pop: Popular ~ Instrument + (Instrument | Subject)
## final.model.dichom.2.voice.pop: Popular ~ Instrument + musician_status_2:Voice + (Instrument |
## final.model.dichom.2.voice.pop:     Subject)
##                                Df   AIC   BIC  logLik deviance  Chisq
## final.model.pop                10 10104 10162 -5042.0    10084
## final.model.dichom.2.voice.pop 15 10107 10194 -5038.3    10077 7.4103
##                                Chi Df Pr(>Chisq)
## final.model.pop
## final.model.dichom.2.voice.pop      5     0.1919
```

It seems that among all the interactions between the dichotomized variable for self declare, the interaction between the dichotomized variable and the harmony is most significant, with a high test statistic and low p-value. The other interactions do not seem to be significant as they have low test statistics and high p-values according to the results of the ANOVA test. This is consistent with the results of ANOVA tests in question 4, when predicting classical music scores.