The Identification of Classical and Popular Music

Kylie Huang

zhiyanh@andrew.cmu.edu

December 6th 2019

Abstract

We try to find how personal experience and music characteristics affect people's decisions on whether a piece of music is classical music or popular music. Based on the data collected by Ivan Jimenez and Vincent Rossi in 2012 at the University of Pittsburgh, we applied exploratory data analysis and multilevel model building techniques to explore the factors that may affect the classical and popular ratings. After the analysis, we found that Instrument has the strongest effect on Classical rating but not on Popular rating. The I-V-VI in Harmony and contrary motion in Voice Leading have the strongest association with classical rating, among all levels in Voice and Harmony. We notice that musicians and non-musicians have different ways to identify classical music. And the factors that affect people's classical and popular ratings are also different. The data we analyze have data missingness problem, so we may need more details about the data collection process to confirm the relationship.

Introduction

The identification of classical and popular music is not an easy task, especially for music that is not well known as popular or classical. However, if you ask a person to choose whether a piece of music is popular music or classical music, you will get an answer. How this person identify the type of this music? Does personal background affect music identification? Are there any characteristics in this piece of music that matches the characteristics of classical music? In this research, we are going to explore the factors that may affect the public's identification of classical or popular music from the following perspectives (Junker, 2019):

- What experimental factor, or combination of factors, has the strongest influence on ratings?
  - Does Instrument exert the strongest influence among the three design factors (Instrument, Harmonic Motion, Voice Leading), as the researchers suspect?
  - Among the levels of Harmonic Motion does I-V-vi have a strong association (the strongest?) 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 a strong (the strongest?) association with classical ratings?
- Are there differences in the way that musicians and non-musicians identify classical music?
- Are there differences in the things that drive classical, vs. popular, ratings?

Methods

The data comes from the Canvas webpage for 36-617 Applied Linear Model at Carnegie Mellon University. Originally, the data was collected by Ivan Jimenez and Vincent Rossi in 2012 to study the factors that assist listeners in identifying the music as popular music or classical music. The researchers presented 36 musical stimuli to 70 undergraduate students at the University of Pittsburgh and received the scores for the following variables:

- Classical = How classical does the stimulus sound?
- Popular = How popular does the stimulus sound?
- Subject = Unique subject ID
- Harmony = Harmonic Motion (4 levels, includes I-V-VI, I-VI-V, I-V-IV, IV-I-V)
- Instrument = Instrument (3 levels, includes String Quartet, Piano, Electric Guitar)
- Voice = Voice Leading (3 levels, includes Contrary Motion, Parallel 3rds, Parallel 5ths)
- Selfdeclare = Are you a musician? (1-6, 1=not at all)
- OMSI = Score on a test of musical knowledge
- X16.minus.17 = Auxiliary measure of listener's ability to distinguish classical vs popular music
- ConsInstr = How much did you concentrate on the instrument while listening (0-5, 0=not at all)
- ConsNotes = How much did you concentrate on the notes while listening? (0-5, 0=not at all)
- Instr.minus.Notes = Difference between prev. two variables
- PachListen = How familiar are you with Pachelbel's Canon in D (0-5, 0=not at all)
- ClsListen = How much do you listen to classical music? (0-5, 0=not at all)
- KnowRob = Have you heard Rob Paravonian's Pachelbel Rant (0-5, 0=not at all)
- KnowAxis = Have you heard Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music? (0-5, 0=not at all)
- X1990s2000s = How much do you listen to pop and rock from the 90's and 2000's? (0-5, 0=not at all)
- X1990s2000s.minus.1960s1970s = Difference between prev variable and a similar variable referring to 60's and 70's pop and rock.
- CollegeMusic = Have you taken music classes in college (0=no, 1=yes)
- NoClass = How many music classes have you taken?
- APTheory = Did you take AP Music Theory class in High School (0=no, 1=yes)
- Composing = Have you done any music composing (0-5, 0=not at all)
- PianoPlay = Do you play piano (0-5, 0=not at all)
- GuitarPlay = Do you play guitar (0-5, 0=not at all)
- X1stInstr = How proficient are you at your first musical instrument (0-5, 0=not at all)
- X2ndInstr = Same, for second musical instrument

Before the analysis, we performed data preprocessing, including deleting missing data and transformed data. The dataset contains 2520 records, but not all records are valid and complete. X1stInstr contains 1512 missing data points, and X2ndInstr contains 2196 missing data points. Both variables were removed because of the large amount of missingness. Other variables also have the same missing data issue, but the amount of missing data points is not as much as X1stInstr and X2sInstr, so only the incomplete records were removed, instead of the variables. Other than the missingness, there are some decimal numbers, zeros, and mistyped data (such as 19) that exist in Classical and Popular, which are not valid and removed for further analysis. After cleaning up the missingness and invalid data points, the dataset for classical rating analysis (the dataset without Popular) includes 1532 records, and the popular rating dataset (the dataset without Classical) includes 1517 records. To satisfy the assumption that data should follow the normal distribution, we performed log transformation on OMSI.

The analysis was focusing on two subjects: classical ratings and popular ratings. For the classical ratings, we constructed a multilevel model with Instrument, Harmonic Motion, and Voice Leading to understand their effects on the ratings. Then we add the effects from other variables into the multilevel model to have more comprehensive views of the association between the factors and classical rating. Among the variables in the model, the Selfdeclare variable was divided into musicians and non-musicians at level 2 to see the reaction difference between musicians and non-musicians on classical rating. For the feature selection process, we applied AIC to select fix effects and fitLMER.fnc function in LMERConvenienceFunctions package to select random effects and reselect fix effects. For the model selection process, we used the scores from ANOVA test to make decisions. The marginal, condition and random residual plots were also used for assumption validation. The same procedure was applied for the popular rating dataset to find the effects of the variables on people's identification of popular music.

Results

1.Exploratory data analysis

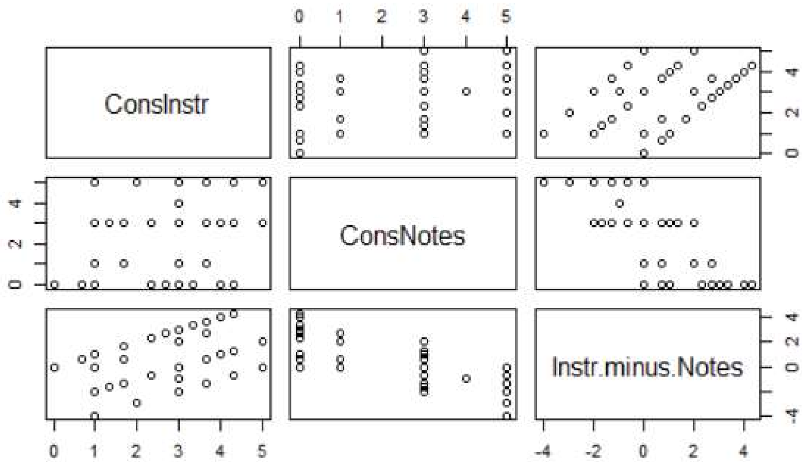


Figure 1: Scatterplot matrix showing the correlation between ConsInstr and Instr.minus.Notes

By looking at the scatterplot matrix of the variables (Appendix Table 2,3,4), we can see the collinearity between variables from the linear patterns. From figure 1, we can see that there is a linear pattern between ConsInstr and Instr.mibus.Notes, and the ConsNotes and Instr.minus.Notes. With the increases of Instr.minus.Notes, the ConsInstr may also increase. At the same time, if the Instr.minus.Notes increases, the ConsNotes is likely to decrease. These patterns show that collinearity exists between ConsInstr and Instr.minus.Notes and ConsNotes. Since we need to satisfy the assumption for the valid model that no collinearity exists between variable, and the instr.minus.notes can always be recovered by minus ConsNotes from ConsInstr, the Instr.minus.Notes is removed.

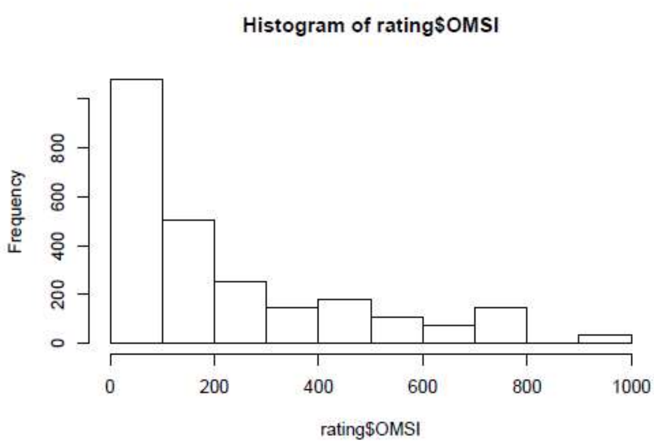


Figure 2: Histogram showing the skewness of OMSI

Other than non-collinearity, normality is also an important assumption for ordinary linear and multilevel models. From figure 2 above, we can see that the distribution of OMSI is skewed to the right, so it is not normally distributed. To solve this issue, log transformation is applied to transform OMSI to satisfy normality.

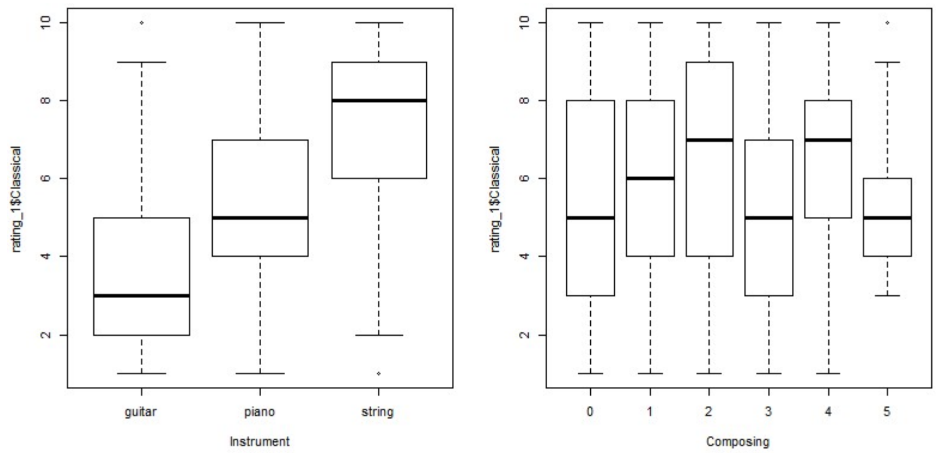


Figure 3: Boxplots showing the effects from different levels of the same variables

By looking at the boxplot matrix of each variable vs. classical rating(Appendix Figure 5), we can roughly detect the effects on the classical rating from each variable. If we look at the figures above (Figure 3), from the plot on the left, we can see that the median of the piano surpasses the upper interquartile range of guitar. It indicates that the effects of Piano and Guitar are different. Also, it is interesting to see that the effects of Guitar, Piano, and String are quite different from each other. However, it is not always the case. If we look at the graph at the right, for the level from 0 to 3, the median and interquartile range have no much difference. But if we compare the effect for level 3,4,5, we can see that the medians are either higher or lower than the interquartile boundaries of the next level. This means that effects from different levels from the same variables can be significantly different, and we may need to pay attention to the levels instead of the variables as a whole in the further analysis.

2.Classical rating

```
Random effects:
 Groups   Name             Variance Std.Dev.
 Subject  Instrumentguitar 0.41794  0.6465
          Instrumentpiano  1.68974  1.2999
          Instrumentstring 2.01951  1.4211
```

```
Random effects:
 Groups   Name             Variance Std.Dev.
 Subject  HarmonyI-IV-V    0.4179   0.6465
          HarmonyI-V-IV    0.7587   0.8711
          HarmonyI-V-VI    1.7053   1.3059
          HarmonyIV-I-V    0.4149   0.6441
```

```
Fixed effects:
                 Estimate Std. Error t value
Instrumentguitar 2.352811   0.871574   2.699
Instrumentpiano  3.953498   0.888718   4.449
Instrumentstring 5.878893   0.894529   6.572
```

```
Fixed effects:
                 Estimate Std. Error t value
HarmonyI-IV-V    2.352813   0.871574   2.700
HarmonyI-V-IV    2.339868   0.876687   2.669
HarmonyI-V-VI    3.200941   0.892416   3.587
HarmonyIV-I-V    2.423816   0.871497   2.781
```

Fixed effects:

	Estimate	Std. Error	t value
Voicecontrary	2.3525485	0.8714648	2.700
Voicepar3rd	1.9667683	0.8711357	2.258
Voicepar5th	2.0497891	0.8712030	2.353

Table 1: Summary table showing the effects of Harmony, Voice, and Instrument on Classical rating

2.1 The influence of instruments on classical rating

We can tell the influence of Harmony, Instrument, and Voice on Classical rating from Table 1 above. In the fixed-effect section, t values are all bigger than 1.96, which is the t statistics for 0.05 p-value in large sample tests. In other words, all levels are significant. Other than the significance, by comparing the coefficients for all the levels, we can see that the string in Instrument has the highest coefficient among all the levels in three variables, which means that String has the strongest influence in three design factors. Other than string, we can also see that the coefficient for piano is the second largest in the levels, which indicates that piano has the second largest influence on classical rating, compare with levels in other variables.

If we combine the random effect and the fixed effect of Instrument, we can see that the significance of Instrument is not deprecated even with the random variance. By minus two times of instrument random effect standard deviation, we can get the lower bound of 95% confidence interval of the effects from Instrument, which are still positive numbers. Therefore, we can conclude that the instrument has the strongest influence on classical rating in three main design factors.

2.2 The association between Harmony I-V-VI and classical rating

	Classical	Harmony_I-IV-V	Harmony_I-V-IV	Harmony_I-V-VI	Harmony_IV-I-V
Classical	1.00000000	-0.04806789	-0.05085003	0.1319529	-0.03271859
Harmony_I-IV-V	-0.04806789	1.00000000	-0.33449477	-0.3327511	-0.33391355
Harmony_I-V-IV	-0.05085003	-0.33449477	1.00000000	-0.3327511	-0.33391355
Harmony_I-V-VI	0.13195294	-0.33275110	-0.33275110	1.0000000	-0.33217290
Harmony_IV-I-V	-0.03271859	-0.33391355	-0.33391355	-0.3321729	1.00000000

Table 2: Correlation between each level in Harmony and Classical

Other than Instrument, Harmony also plays significant roles in classical rating, which can be detected in Table 2. We can see that the coefficient for Harmony I-V-VI is 3.2, which is the largest coefficient, compared with other levels in Harmony. In other words, the Harmony I-V-VI has the largest effect on Classical rating in all the levels in Harmony. At the same time, in the correlation matrix (Table 2), we can see that Classical rating has the largest absolute correlation with Harmony I-V-IV, which means that Harmony I-V-VI has the strongest relationship with Classical rating. If we combine the correlation and the fix effect coefficients, we can see that the Harmony I-V-VI has the strongest relationship and exert the largest effect on Classical rating in all four levels in Harmony.

	Estimate	Std. Error	t value
KnowAxis1:HarmonyI-V-IV	1.4470425	0.8249079	1.754
KnowAxis5:HarmonyI-V-IV	-0.1538648	0.2907567	-0.529
KnowAxis1:HarmonyI-V-VI	-1.4875917	1.3104638	-1.135
KnowAxis5:HarmonyI-V-VI	-1.1018326	0.4616157	-2.387
KnowAxis1:HarmonyIV-I-V	0.6147048	0.8633461	0.712
KnowAxis5:HarmonyIV-I-V	-0.3489073	0.3040657	-1.147
HarmonyI-V-IV:KnowRob1	0.0404018	0.4047802	0.100
HarmonyI-V-VI:KnowRob1	0.9296494	0.6473257	1.436
HarmonyIV-I-V:KnowRob1	0.2286109	0.4242994	0.539
HarmonyI-V-IV:KnowRob5	-0.3986573	0.3670339	-1.086
HarmonyI-V-VI:KnowRob5	1.9027492	0.5786195	3.288
HarmonyIV-I-V:KnowRob5	0.1014195	0.3818915	0.266

Table 3: Summary table showing the interaction of KnowAxis and Harmony, KnowRob and Harmony.

We can see the influence of the interaction between Harmony I-V-VI and KnowRob or KnowAxis from Table 3 above (full summary table in Table 8). We can see that only the t value of the interaction of KnowAxis5 and Harmony I-V-VI and the interaction of KnowRob5 and Harmony I-V-VI have absolute t value that bigger than 1.96, which means only these two levels have significant interaction with harmony I-V-VI. In another words, only in the case that the respondent is very familiar with Pachelbel rants and Comedy bits will be matter on the association between Harmony I-V-VI and Classical rating.

2.3 The association between Voice and Classical

	Classical	Voice_contrary	Voice_par3rd	Voice_par5th
Classical	1.00000000	0.06075679	-0.04123251	-0.01954448
Voice_contrary	0.06075679	1.00000000	-0.49975508	-0.50048972
Voice_par3rd	-0.04123251	-0.49975508	1.00000000	-0.49975508
Voice_par5th	-0.01954448	-0.50048972	-0.49975508	1.00000000

Table 4: Correlation table showing the correlation between each level of Voice and Classical

After understanding the effects of Harmony, now we are turning to the effects of Voice, which are shown in Table 1. We can see that all the levels in Voice are significant because their t values are bigger than 1.96. Also, the voice contrary has the largest coefficient, which means that it has the strongest contribution on the respondent to give higher classical ratings. Coefficients can not show the association of Voice contrary and the classical rating alone. We need a correlation to supplement the conclusion, which is shown in Table 4. From this table, we can see that the correlation between Voice contrary and Classical is 0.06, which is the largest absolute correlation in all levels in Voice. Combining the effects of coefficients and correlation, we can say that the Voice contrary has the strongest association with classical rating in all Voice levels.

2.4 Musicians vs. Non-musicians

	Estimate	Std. Error	t value
HarmonyI-V-IV:MusicianTrue	7.649e-02	2.896e-01	0.264
HarmonyI-V-VI:MusicianTrue	1.501e+00	4.320e-01	3.475
HarmonyIV-I-V:MusicianTrue	2.979e-01	2.823e-01	1.055
Instrumentpiano:MusicianTrue	-5.723e-01	4.653e-01	-1.230
Instrumentstring:MusicianTrue	-1.080e+00	5.859e-01	-1.843
ConsInstr1:MusicianTrue	1.085e+01	1.497e+00	7.248
ConsInstr1.67:MusicianTrue	1.314e+01	1.727e+00	7.607
ConsInstr3:MusicianTrue	-2.354e+00	6.720e-01	-3.503
ConsInstr3.67:MusicianTrue	5.746e-01	5.254e-01	1.094
ConsInstr4.33:MusicianTrue	2.168e+01	2.478e+00	8.751
ClsListen1:MusicianTrue	-1.998e+00	8.419e-01	-2.373
ClsListen3:MusicianTrue	-5.085e+00	9.096e-01	-5.590
ClsListen5:MusicianTrue	-1.693e+00	1.564e+00	-1.082
KnowAxis5:MusicianTrue	4.266e+00	7.882e-01	5.412
KnowRob1:MusicianTrue	-2.447e+01	2.651e+00	-9.230
KnowRob5:MusicianTrue	-1.583e+01	1.760e+00	-8.991
HarmonyI-V-IV:KnowAxis5:MusicianTrue	-3.880e-01	5.401e-01	-0.718
HarmonyI-V-VI:KnowAxis5:MusicianTrue	-1.579e+00	7.806e-01	-2.023
HarmonyIV-I-V:KnowAxis5:MusicianTrue	-9.954e-01	5.510e-01	-1.807

Table 5: Summary table showing the interaction of the variables and Musicians

Another thing that we interested in is whether musicians and non-musicians have different approaches to identify classical music. From Table 5, we can see how musicians and non-musicians interact with the variables differently. (The full summary table in Appendix Table 6) The model includes the two-way interactions of Harmony, Instrument, ConsInstr, ClsListen, KnowAxis, KnowRob, and Musicians, as well as the three-way interaction of Harmony, KnowAxis, and Musician. This means that the musicians and non-musician will bring different effects on the relationship between instrument choice, harmony characteristics, concentration on instruments, frequency of listening to classical music, composing level, the familiarity of Pachelbel rants, the familiarity of Comedy bits, the interaction between Harmony characteristics and familiarity of Pachelbel, and Classical music. For example, compared with the music played in harmony I-IV-V, musicians may give 1.501 more than non-musicians in classical rating for the music play in harmony I-V-VI. Other than the effects on other variables, whether the respondent is a musician can affect their identification on classical music directly. The classical rating given by musicians increases by 2.45, compared with the people who are not musicians.

3. Popular rating

3.1 Influence from Voice, Instrument and Harmony on Popular ratings

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	Instrumentguitar	0.5746	0.7581
	Instrumentpiano	2.2172	1.4890
	Instrumentstring	1.5017	1.2254

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	HarmonyI-IV-V	0.5746	0.7581
	HarmonyI-V-IV	1.1031	1.0503
	HarmonyI-V-VI	1.4101	1.1875
	HarmonyIV-I-V	0.8232	0.9073

Fixed effects:

	Estimate	Std. Error	t value
HarmonyI-IV-V	7.92085	0.72401	10.940
HarmonyI-V-IV	7.93137	0.73241	10.829
HarmonyI-V-VI	7.63725	0.73723	10.359
HarmonyIV-I-V	7.63572	0.72796	10.489

```
Fixed effects:
                  Estimate Std. Error t value
Voicecontrary      7.92085    0.72401  10.940
Voicepar3rd        8.08376    0.72410  11.164
Voicepar5th        8.13410    0.72409  11.234

Fixed effects:
                  Estimate Std. Error t value
Instrumentguitar  7.96485    0.71911  11.076
Instrumentpiano   6.80873    0.74514   9.138
Instrumentstring  4.99798    0.73387   6.810
```

Table 6: Summary table showing the effects of three main factors on Popular ratings

After analyzing the effects on classical rating, now we are turning to the analysis on popular ratings. Table 6 shows the effects of Harmony, Instrument, and Voice on Popular rating. From this figure, we can see that all the fixed effects levels in three main design factors are significant because their t values are bigger than 1.96. Also, we know that VoicePar3rd and VoicePar5th are two levels that have the biggest coefficients. The last level in Voice, the Contrary, is also bigger than most of the other levels. So, Voice has the strongest influence on Popular rating as a fixed effect, instead of Instrument. If we turn to the random effects, we can see that the random effect of Voice is the only one in the three main design factors that no significant enough to be included in the model. In other words, the effects of Voice are not significantly different among the respondents. Combining the random effects and fixed effects, we can say that voice has the strongest effects on Popular rating, and its effects are generally agreed by the respondents.

4. Popular ratings vs. Classical ratings

```
Popular ~ ConsInstr + PachListen + KnowAxis + GuitarPlay + Voice +
    Instrument + ConsNotes + KnowRob + Harmony + (1 | Subject) +
    (Instrument | Subject) + (Harmony | Subject)

Random effects:
 Groups    Name             Variance Std.Dev.
 Subject   (Intercept)      0.01021  0.1010
 Subject.1 (Intercept)      1.01990  1.0099
           Instrumentpiano  1.75140  1.3234
           Instrumentstring 2.64730  1.6271
 Subject.2 (Intercept)      0.14833  0.3851
           HarmonyI-V-IV    0.12660  0.3558
           HarmonyI-V-VI    0.76362  0.8739
           HarmonyIV-I-V    0.34898  0.5907
 Residual                   2.57395  1.6044
Number of obs: 1517, groups:  Subject, 43

Fixed effects:
               Estimate Std. Error t value  GuitarPlay1      -0.81378    0.85544  -0.951
(Intercept)     7.46358    1.03386   7.219  GuitarPlay2       1.18933    0.85907   1.384
ConsInstr0.67  -3.48152    1.35108  -2.577  GuitarPlay4       0.79224    0.87818   0.902
ConsInstr1      1.69910    0.89199   1.905  GuitarPlay5      -0.07577    0.45986  -0.165
ConsInstr1.67   1.28519    0.99193   1.296  Voicepar3rd       0.16453    0.10104   1.628
ConsInstr2.33   0.69287    0.83988   0.825  Voicepar5th       0.21451    0.10104   2.123
ConsInstr2.67  -0.94199    1.00355  -0.939  Instrumentpiano  -1.15680    0.22555  -5.129
ConsInstr3     -0.68920    0.94064  -0.733  Instrumentstring -2.96728    0.26859 -11.048
ConsInstr3.33   2.35037    1.18174   1.989  ConsNotes1       -0.33607    0.68056  -0.494
ConsInstr3.67   0.02855    1.05417   0.027  ConsNotes3       -0.33688    0.56969  -0.591
ConsInstr4      0.18606    1.10744   0.168  ConsNotes4       -0.22790    0.94064  -0.242
ConsInstr4.33   0.18593    0.94565   0.197  ConsNotes5        0.13204    0.65062   0.203
ConsInstr5      0.69422    0.97360   0.713  KnowRob1          0.82899    0.51548   1.608
PachListen3     0.33120    0.95404   0.347  KnowRob5         -0.49825    0.48736  -1.022
PachListen5    -1.03003    0.73219  -1.407  HarmonyI-V-IV     0.01093    0.12875   0.085
KnowAxis1      -0.62116    1.22188  -0.508  HarmonyI-V-VI    -0.28356    0.17718  -1.600
KnowAxis5       0.50489    0.47375   1.066  HarmonyIV-I-V    -0.28400    0.14711  -1.930
```

Table 7: Summary table showing the final classical rating model

Other than the three main effects, the final model (Table 7)for Popular ratings also includes fix effects from ConsInstr, PachListen, KnowAxis, GuitarPlay, ConsNotes, KnowRob. From this model, we can see that ConsInstr0.67, ConsInstr3.33, VoicePar5th, InstrumentPiano, InstrmentString are significant. If we hold other variables constant, the Popular rating decreases 3.48, if the ConsInstr scores increase from 0 to 0.67. The Popular rating increases 2.35, if the ConsInstr score increases from 0 to 3.33. The Popular rating decrease 1.16, if the instrument that plays the music switches from guitar to piano. The Popular rating decreases 2.97, if the instrument that plays the music switches from guitar to string.

Significance is not the only thing we can learn from the Popular rating model. If we compare the coefficients for each variable, we can see that the respondents whose ConsInstr score is 3.33 give the highest Popular rating, compared with all other levels in ConsInstr; the respondents who score 3 in PacheListen give the highest Popular rating, compared with all other levels in PachListen; the respondents who score 5 in KnowAxis give the highest Popular rating, compared with all other levels in KnowAxis; the respondents who score 2 in GuitarPlay give the highest Popular rating, compared with all other levels in GuitarPlay; the respondent who scores 5 in ConsNotes give the highest Popular rating, compare with all other levels in ConsNotes; the respondent who scores 1 in KnowRob give the highest Popular rating, compare with all other levels in KnowRob; the music song in Voice 5^{th} receive the highest Popular rating, compared with all other levels in Voice; the music played in Guitar received the highest Popular rating, compared with all other levels in Instrument; the music played in Harmony I-V-IV received the highest Popular rating, compared with all other levels in Harmony.

```
Formula: Classical ~ KnowAxis:Harmony + KnowRob:Harmony + Voice + Harmony +
    Instrument + Selfdeclare + ConsInstr + PachListen + ClsListen +
    Composing + PianoPlay + GuitarPlay + KnowAxis + KnowRob +
    (1 + Instrument + Harmony | Subject)

Random effects:
 Groups   Name            Variance Std.Dev. Corr
 Subject  (Intercept)     0.41794  0.6465
          Instrumentpiano 2.05391  1.4331   -0.42
          Instrumentstring 3.55856 1.8864   -0.80  0.61
          HarmonyI-V-IV   0.08362  0.2892    0.69 -0.65 -0.96
          HarmonyI-V-VI   1.26741  1.1258    0.01 -0.39 -0.60  0.71
          HarmonyIV-I-V   0.06848  0.2617   -0.21 -0.12 -0.11  0.00  0.38
 Residual                 2.40070  1.5494
Number of obs: 1532, groups:  Subject, 43

Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	2.3956209	0.8728693	2.745
Voicepar3rd	-0.3856678	0.0970333	-3.975
Voicepar5th	-0.3027265	0.0969731	-3.122
HarmonyI-V-IV	0.0520839	0.1445788	0.360
HarmonyI-V-VI	0.7710711	0.2415818	3.192
HarmonyIV-I-V	0.1082178	0.1457798	0.742
Instrumentpiano	1.6007023	0.2393616	6.687
Instrumentstring	3.5271237	0.3036268	11.617
Selfdeclare2	-1.1280783	0.4357443	-2.589
Selfdeclare3	-1.7979210	0.4775014	-3.765
Selfdeclare4	-0.2354878	0.6193972	-0.380
Selfdeclare5	-6.8355375	0.9674489	-7.066
Selfdeclare6	-2.3868498	1.2161393	-1.963
ConsInstr0.67	1.9858815	0.9075098	2.188
ConsInstr1	-1.8216574	0.5783951	-3.150
ConsInstr1.67	0.0618579	0.4766702	0.130
ConsInstr2.33	0.0568422	0.4083446	0.139
ConsInstr2.67	1.5470102	0.4576518	3.380
ConsInstr3	-0.6093221	0.5176882	-1.177
ConsInstr3.33	-1.9941695	0.6808136	-2.929
ConsInstr3.67	-0.3228774	0.4843840	-0.667
ConsInstr4	1.4812705	0.6334952	2.338
ConsInstr4.33	0.5047918	0.4348764	1.161
ConsInstr5	0.3140768	0.4163215	0.754
PachListen3	0.7305724	0.4808619	1.519
PachListen5	1.9183668	0.3928462	4.883
ClsListen1	0.1019740	0.3308936	0.308
ClsListen3	0.7018126	0.3790424	1.852
ClsListen4	3.1058553	0.8415697	3.691
ClsListen5	-1.9267916	0.5310089	-3.629
Composing1	0.7839113	0.3148308	2.490
Composing2	-0.0005436	0.5263683	-0.001
Composing3	-2.6275399	0.5622118	-4.674
Composing4	2.0127585	0.3845061	5.235
PianoPlay1	0.8332162	0.2789695	2.987
PianoPlay4	3.5148028	0.5367585	6.548
PianoPlay5	1.2863735	0.5480561	2.347
GuitarPlay1	4.4623354	0.7162266	6.230
GuitarPlay2	-2.5456498	0.9443145	-2.696
GuitarPlay4	2.3398704	1.0192611	2.296
GuitarPlay5	0.0586703	0.4923465	0.119
KnowAxis1	-2.3364115	1.7260995	-1.354
KnowAxis5	0.1834732	0.3559355	0.515
KnowRob1	2.1685322	0.6835156	3.173
KnowRob5	-0.5357766	0.4497033	-1.191
KnowAxis1:HarmonyI-V-IV	1.4470425	0.8249079	1.754
KnowAxis5:HarmonyI-V-IV	-0.1538648	0.2907567	-0.529
KnowAxis1:HarmonyI-V-VI	-1.4875917	1.3104638	-1.135
KnowAxis5:HarmonyI-V-VI	-1.1018326	0.4616157	-2.387
KnowAxis1:HarmonyIV-I-V	0.6147048	0.8633461	0.712
KnowAxis5:HarmonyIV-I-V	-0.3489073	0.3040657	-1.147
HarmonyI-V-IV:KnowRob1	0.0404018	0.4047802	0.100
HarmonyI-V-VI:KnowRob1	0.9296494	0.6473257	1.436
HarmonyIV-I-V:KnowRob1	0.2286109	0.4242994	0.539
HarmonyI-V-IV:KnowRob5	-0.3986573	0.3670339	-1.086
HarmonyI-V-VI:KnowRob5	1.9027492	0.5786195	3.288
HarmonyIV-I-V:KnowRob5	0.1014195	0.3818915	0.266

Table 8: Summary table showing the final popular rating model

The final model for Classical rating is quite different from the Popular rating model(Table 8). The Classical rating model contains Voice, Harmony, Instrument, Selfdeclare, ConsInstr, PacheListen, ClsListen, Composing, PianoPlay, GuitarPlay, KnowAxis, KnowRob, the interaction between KnowAxis and Harmony, the interaction between Harmony and KnowRob and the random effects from Instrument and Harmony. In this model, other than the two levels in Instruments, ConsInstr0.67, ConsInstr3.33 that have significant effects on Popular rating, the VoicePar3rd, Voicepar5th, Harmony I-V-VI, Selfdeclare2, Selfdeclare3, Selfdeclare5, Selfdeclare6, ConsInstr2.67,3,4, PachListen5, ClsListen4,5, Composing1,3,4,PianoPlay1,4,5, GuitarPlay1,2,4, KnowRob5, the interaction between KnowAxis5 and Harmony I-V-IV, and the interaction between Harmony I-V-VI and KnowRob5 is also significant when people identify Classical music.

It is interesting to see that there are some variables that both exist in Classical rating model and Popular rating model besides three main design factors, such as ConsInstr, PachListen, KnowAxis, GuitarPlay, and KnowRob, which means that these factors are important for respondents to identify the music type in general, not matter classical or popular music. Other than that, if we compare the signs of the coefficients of the shared variables, we can see that most of the signs in popular models are opposite to the ones in the classical rating model. For example, the sign for ConsInstr 0.67 in the Classical rating model is positive while in Popular rating model is negative. This implies that the factors that enhance the possibility of the music to be classical music may also strengthen the possibility of the music to be popular music in people's minds.

Discussion

In general, if we only compared the random and fixed effects brought by instrument, harmony, and voice, what instruments the stimuli are played with have the most significant effects on classical rating. By checking the coefficients and the correlation between all levels in Voice and Classical rating, and the correlation between all levels in Harmony and Classical rating, we know that I-V-VI in Harmonic Motion and contrary in Voice Leading have the strongest association with classical ratings. Other than that, by comparing the interaction for musicians and the variables in the classical rating model, we can see that musicians interpret the relationship between Harmony, Instrument, ConsInstr, ClsListen, KnowAxis, KnowRob, and the interaction between Harmony and KnowAxis differently with non-musicians. If we compare the effects of the factors on Popular rating and Classical rating, we can see that the variables that have influence on Classical rating more than the variables for Popular rating. Also, for the variables that have influence on both classical rating and popular rating, the signs are opposite.

There are some interesting points in the comparison of the popular rating model and classical rating model. The signs of the instrument, harmony and voice coefficients in popular rating mode and the signs in the classical rating model are flipped. It is reasonable because there are only two options for respondents, either they think this music is classical music or popular music. For some variables that people think are characteristics of classical music, they may also think the existence of these classical music characters may decrease the chance of the music being popular music. Therefore, the signs of the variables that exist in the popular rating model are opposite from the signs of the same variables in the classical rating model.

Another interesting point is that Voice is the only main design variable that does not have significant random effects on Classical rating and Popular rating. The reason maybe that people may have diverse ideas about what kind of harmony or instrument should be used in popular or popular music, but they have general agreement on what kind of Voice Leading should show up in popular music. For example, some respondents may disagree that popular music should be played by guitar but not piano, while most of the respondents agree that popular music should be sung in 5th for Voice Leading.

If we extend our comparison from the effects of three design factors to other variables in the model, we can see that the classical model includes a lot more variables in the popular rating model. The classical model includes the musician scores, classical music listening frequency, piano practice level and composing level, which are not in the popular rating model. Among these variables, we can see the relationship between classical music listening frequency, piano practice levels and classical music by their definitions. Also, since musician scores and composing level represents the level of music background, which may be necessary for classical music appreciation, it is reasonable that these are included in the classical rating level. On the other hand, popular music is easy to enjoy for everyone, even the one without prior music experience, so the musician level, composing level, classical music listening frequency, and piano practice level may not be important.

The missing and invalid data in the dataset brings some limitations to the models. Classical and Popular contain some decimal numbers while the definition indicates that the data should be 1-5 integers. Also, there are 1512 missing data points in X1stInstr and 2196 missing data points in X2ndInstr. These two variables are removed because they only contain about1/2 or 1/3 of valid data points. However, these two variables may contain some information that brings significant effects on classical rating or popular rating. Therefore, the model we built may not represent how people identify classical and popular music completely and accurately.

Another limitation is that we do not know how these data are collected, which may bring some unnecessary noises into the analysis. For example, some people may not hear the music clearly, or they are impatient to fill up the questions because they are tired. All these situations may bring some noise into our models and may affect model accuracy.

In the future, we may want to talk with Ivan Jimenez and Vincent Rossi about the data collection methods to figure out the reasons behind the missing data points and do imputation. Also, to have more comprehensive ideas about the association between the factors and type of music, we can sample the students by years at college and build a new model with more comprehensive data.

Reference

Jimenez, I., &Rossi, V., (2013), *The Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music*[PowerPoint slides], Retrieved fromhttps://canvas.cmu.edu/courses/11853/files/folder/hw10/presentation?preview=4203515.

Junker, B.(2019), *Project 3: Do You Hear Classical Or Popular Music?* , Retrieved from https://canvas.cmu.edu/courses/11853/files/folder/project03?preview=4236448

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

Appendix

```
knitr::opts_chunk$set(echo = TRUE)
library(MASS)
library(lme4)
library(LMERConvenienceFunctions)
library(dplyr)
library(arm)
library(RLRsim)
library(ggplot2)
rating = read.csv("ratings.csv")
# filter our first 12
rating = rating[c(2:25, 27, 28)]
```

EDA

```r
# data summary
summary(rating)
```

```
##     Subject        Harmony      Instrument        Voice        Selfdeclare   
##  15      :  36   I-IV-V:630   guitar:840   contrary:840   Min.   :1.000  
##  16      :  36   I-V-IV:630   piano :840   par3rd  :840   1st Qu.:2.000  
##  17      :  36   I-V-VI:630   string:840   par5th  :840   Median :2.000  
##  18b     :  36   IV-I-V:630                               Mean   :2.443  
##  19      :  36                                            3rd Qu.:3.000  
##  20      :  36                                            Max.   :6.000  
##  (Other):2304                                                            
##       OMSI        X16.minus.17      ConsInstr       ConsNotes    
##  Min.   : 11.0   Min.   :-4.000   Min.   :0.000   Min.   :0.000  
##  1st Qu.: 49.0   1st Qu.: 0.000   1st Qu.:1.670   1st Qu.:0.750  
##  Median :145.5   Median : 1.000   Median :3.000   Median :3.000  
##  Mean   :225.9   Mean   : 1.721   Mean   :2.857   Mean   :2.533  
##  3rd Qu.:323.0   3rd Qu.: 3.000   3rd Qu.:4.330   3rd Qu.:5.000  
##  Max.   :970.0   Max.   : 9.000   Max.   :5.000   Max.   :5.000  
##                                                   NA's   :360    
##  Instr.minus.Notes   PachListen      ClsListen        KnowRob      
##  Min.   :-4.0000   Min.   :0.000   Min.   :0.000   Min.   :0.0000  
##  1st Qu.: 0.0000   1st Qu.:5.000   1st Qu.:1.000   1st Qu.:0.0000  
##  Median : 0.3350   Median :5.000   Median :3.000   Median :0.0000  
##  Mean   : 0.6857   Mean   :4.515   Mean   :2.159   Mean   :0.7692  
##  3rd Qu.: 2.0000   3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:0.0000  
##  Max.   : 4.3300   Max.   :5.000   Max.   :5.000   Max.   :5.0000  
##                    NA's   :72      NA's   :36      NA's   :180     
##     KnowAxis       X1990s2000s    X1990s2000s.minus.1960s1970s
##  Min.   :0.0000   Min.   :0.000   Min.   :-4.000              
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.: 0.000              
##  Median :0.0000   Median :5.000   Median : 2.000              
##  Mean   :0.9032   Mean   :4.061   Mean   : 2.015              
##  3rd Qu.:0.0000   3rd Qu.:5.000   3rd Qu.: 3.000              
##  Max.   :5.0000   Max.   :5.000   Max.   : 5.000              
##  NA's   :288      NA's   :144     NA's   :180                 
##   CollegeMusic      NoClass          APTheory        Composing
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0  
##  1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0  
##  Median :1.000   Median :1.0000   Median :0.0000   Median :0  
##  Mean   :0.791   Mean   :0.9194   Mean   :0.2344   Mean   :1  
##  3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:2  
##  Max.   :1.000   Max.   :8.0000   Max.   :1.0000   Max.   :5  
##  NA's   :108     NA's   :288      NA's   :216      NA's   :72 
##    PianoPlay       GuitarPlay       X1stInstr       X2ndInstr    
##  Min.   :0.000   Min.   :0.0000   Min.   :1.000   Min.   :0.000  
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000  
##  Median :0.000   Median :0.0000   Median :3.500   Median :1.000  
##  Mean   :1.086   Mean   :0.6857   Mean   :2.786   Mean   :1.556  
##  3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:2.000  
```

```
##  Max.   :5.000   Max.   :5.0000   Max.   :5.000   Max.   :4.000
##                                   NA's   :1512    NA's   :2196
##    Classical         Popular
##  Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 4.000   1st Qu.: 4.000
##  Median : 6.000   Median : 5.000
##  Mean   : 5.783   Mean   : 5.381
##  3rd Qu.: 8.000   3rd Qu.: 7.000
##  Max.   :19.000   Max.   :19.000
##  NA's   :27       NA's   :27
```

```
# check the unique values in dataset
sample = sapply(rating,unique)
lapply(sample,sort)
```

```
## $Subject
##  [1] 15   16   17   18b  19   20   21   22   23   24   25   26   28   29
## [15] 30   31   32   33   34   35   36   37   38   39   40   41   42   43
## [29] 44.1 44.2 45   46   47   48   49   50   51   52   53   54   55   56
## [43] 57   58   59   60   61   62   63   64   65   66   70   71   73   74
## [57] 75   76   77   78   79   80   81   82   83   90   91   93   94   98
## 70 Levels: 15 16 17 18b 19 20 21 22 23 24 25 26 28 29 30 31 32 33 ... 98
##
## $Harmony
## [1] I-IV-V I-V-IV I-V-VI IV-I-V
## Levels: I-IV-V I-V-IV I-V-VI IV-I-V
##
## $Instrument
## [1] guitar piano  string
## Levels: guitar piano string
##
## $Voice
## [1] contrary par3rd   par5th
## Levels: contrary par3rd par5th
##
## $Selfdeclare
## [1] 1 2 3 4 5 6
##
## $OMSI
##  [1]  11  14  15  18  20  21  23  29  30  31  38  40  44  46  49  55  67
## [18]  68  82  88  94  96  97 122 127 142 145 146 147 150 164 179 180 194
## [35] 199 201 204 233 234 259 277 319 323 325 345 421 425 466 481 482 541
## [52] 567 586 642 649 734 749 759 784 970
##
## $X16.minus.17
##  [1] -4.0 -2.0 -1.0 -0.5  0.0  1.0  2.0  3.0  4.0  5.0  6.0  7.0  9.0
##
## $ConsInstr
##  [1] 0.00 0.67 1.00 1.33 1.67 2.00 2.33 2.67 3.00 3.33 3.67 4.00 4.33 5.00
##
```

```
## $ConsNotes
## [1] 0 1 3 4 5
##
## $Instr.minus.Notes
##  [1] -4.00 -3.00 -2.00 -1.67 -1.33 -1.00 -0.67  0.00  0.67  1.00  1.33
## [12]  1.67  2.00  2.33  2.67  3.00  3.33  3.67  4.00  4.33
##
## $PachListen
## [1] 0 1 2 3 4 5
##
## $ClsListen
## [1] 0 1 3 4 5
##
## $KnowRob
## [1] 0 1 5
##
## $KnowAxis
## [1] 0 1 5
##
## $X1990s2000s
## [1] 0 2 3 4 5
##
## $X1990s2000s.minus.1960s1970s
## [1] -4 -3 -2  0  1  2  3  4  5
##
## $CollegeMusic
## [1] 0 1
##
## $NoClass
## [1] 0 1 2 3 4 8
##
## $APTheory
## [1] 0 1
##
## $Composing
## [1] 0 1 2 3 4 5
##
## $PianoPlay
## [1] 0 1 2 4 5
##
## $GuitarPlay
## [1] 0 1 2 4 5
##
## $X1stInstr
## [1] 1 2 3 4 5
##
## $X2ndInstr
## [1] 0 1 2 3 4
##
## $Classical
```

```
## [1] 0.0 1.0 2.0 3.0 3.5 4.0 4.2 4.6 5.0 6.0 7.0 8.0 9.0 9.5
## [15] 10.0 19.0
##
## $Popular
## [1] 0.0 1.0 2.0 3.0 3.5 4.0 4.2 4.6 5.0 6.0 6.8 7.0 8.0 9.0
## [15] 10.0 19.0
```

```r
# Check the normality of continous variables and do transformation
par(mfrow = c(2,3))
hist(rating$Classical)
hist(rating$Popular)
hist(rating$OMSI)
hist(rating$X16.minus.17)
hist(rating$Instr.minus.Notes)
rating$OMSI = log(rating$OMSI)

# Check correlation between variables
pairs(rating[,c(1:10)])
```

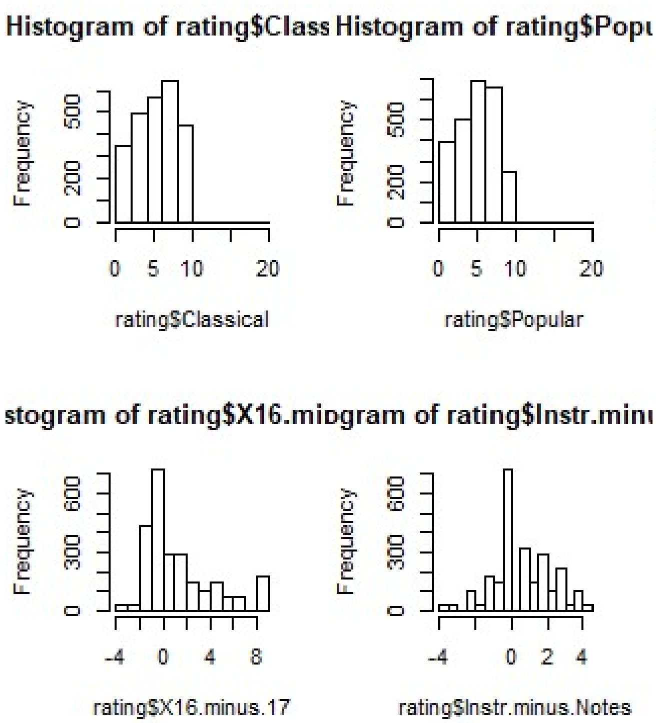


Figure 1

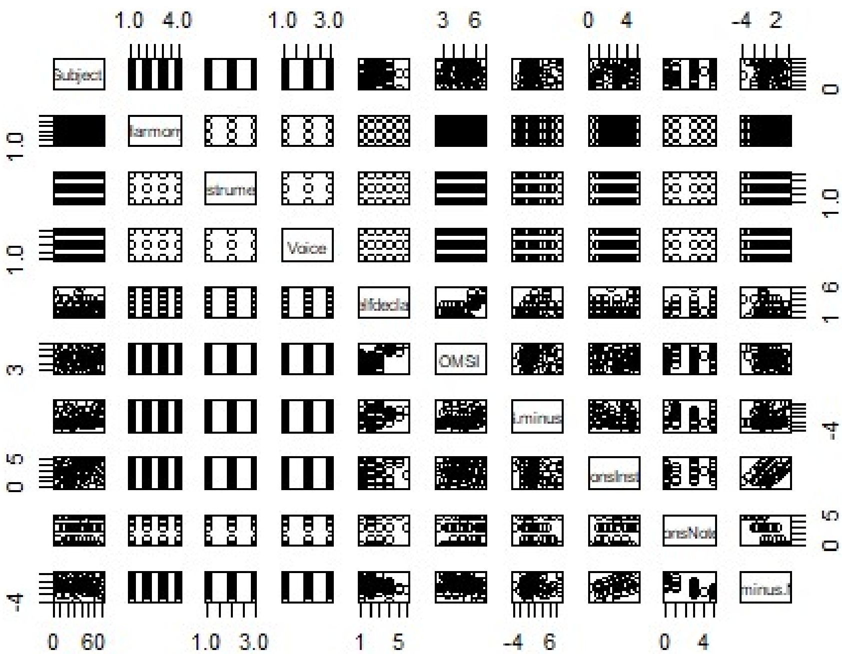


Figure 2

```r
pairs(rating[,c(10:20)])
```

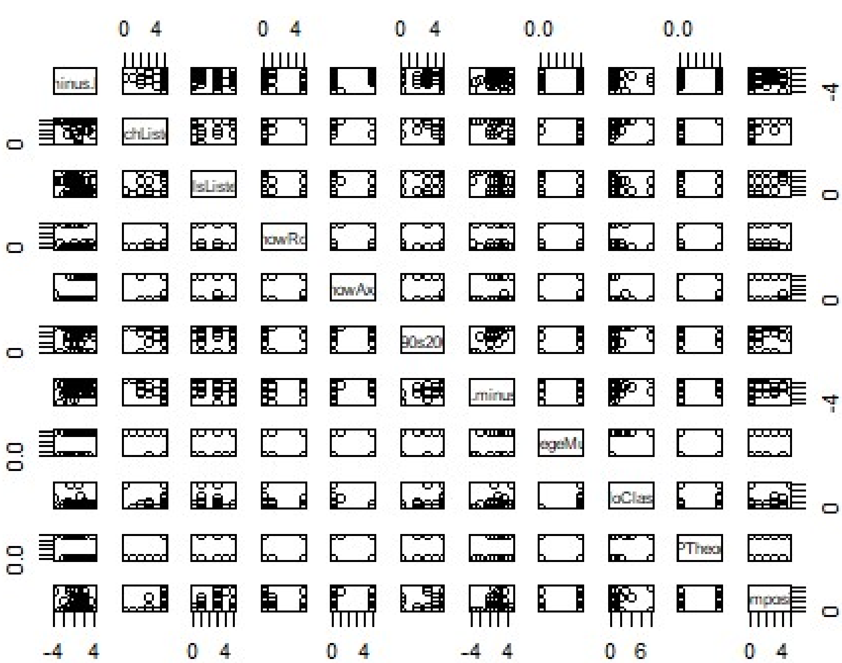


Figure 3

```r
pairs(rating[,c(20:23)])
```

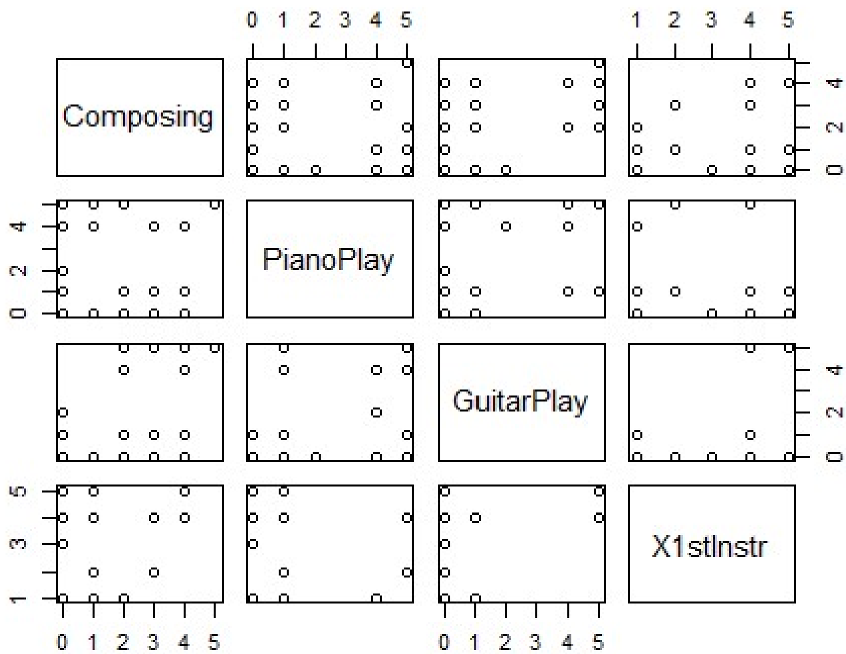
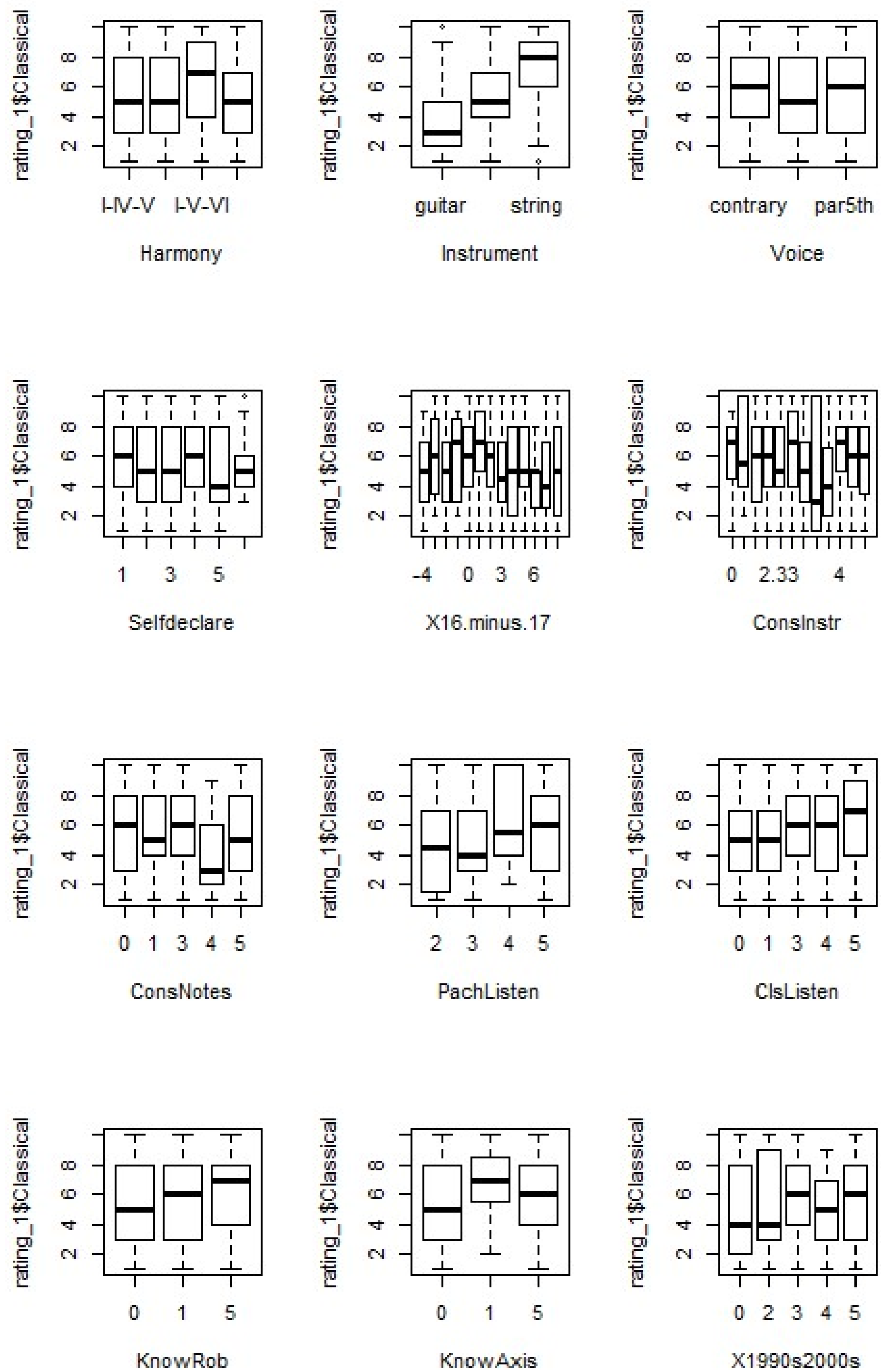


Figure 4

Classical model

```r
# Remove x1instr and x2instr and Popular
rating_1 = rating[c(1:22,25)]

# Drop Instr.minus.Notes, because the distribution is not normal and it is hi
ghly correlated with ConsNote
rating_1 = rating_1[c(1:9,11:23)]

# Drop the Classicle rate == 0, NA, decimal numbers and change 19 to 10
rating_1 = rating_1 %>% filter(rating_1$Classical !=0, !is.na(rating_1$Classi
cal), rating_1$Classical != 3.5, rating_1$Classical != 4.2, rating_1$Classica
l != 4.6, rating_1$Classical != 9.5)
rating_1 = rating_1 %>% mutate(Classical=replace(Classical, Classical==19, 1
0))

# Get rid of the rest of NAs
rating_1 = na.omit(rating_1)

# Do boxplot and see the relation between classical rating an variables
par(mfrow = c(2,3))
```

```
for (i in names(rating_1%>%dplyr::select(-OMSI,-Classical, - Subject))) {boxplot(rating_1$Classical ~ rating_1[,i], xlab = i)}
```



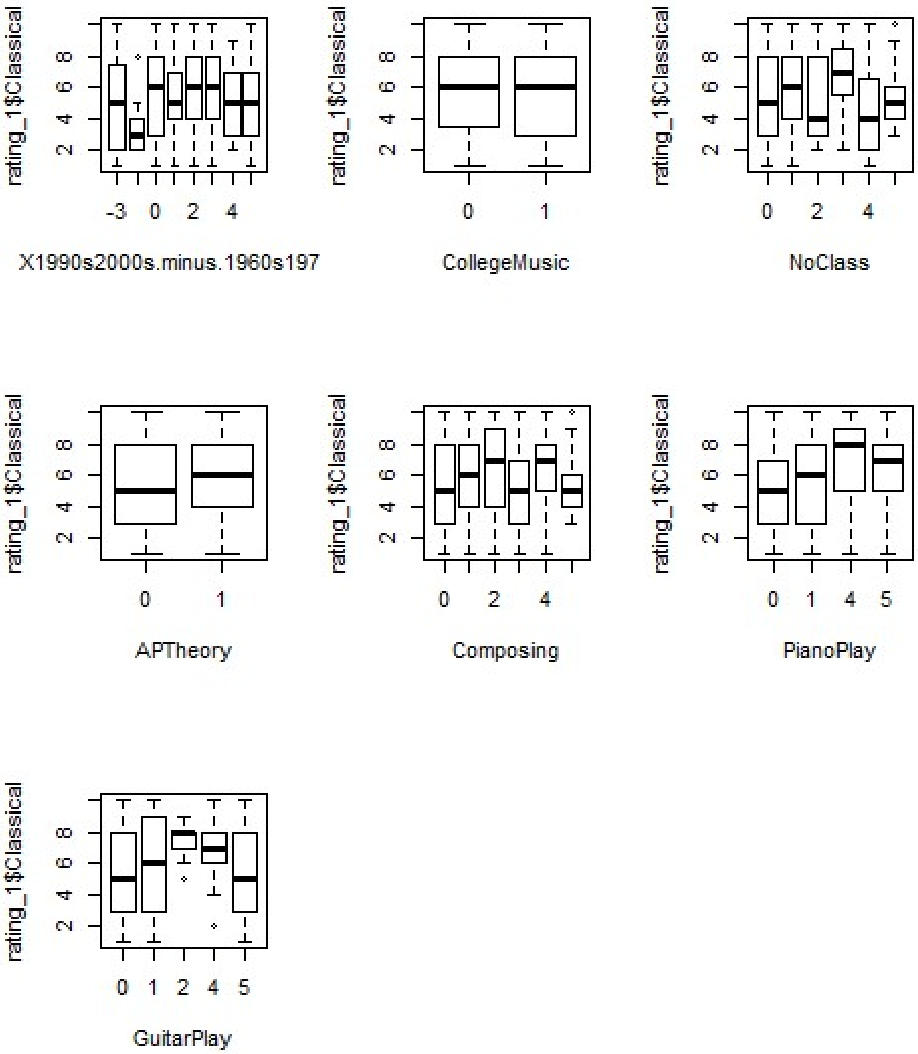

rating_1$Classical
X1990s2000s.minus.1960s197
CollegeMusic
NoClass
APTheory
Composing
PianoPlay
GuitarPlay

Figure 5

Classsical rating final model

```
# Check the type of variables and change some variables to factors
sapply(rating_1, class)
rating_1$Selfdeclare = as.factor(rating_1$Selfdeclare)
rating_1$ConsInstr = as.factor(rating_1$ConsInstr)
rating_1$ConsNotes = as.factor(rating_1$ConsNotes)
rating_1$PachListen = as.factor(rating_1$PachListen)
rating_1$ClsListen = as.factor(rating_1$ClsListen)
rating_1$KnowRob = as.factor(rating_1$KnowRob)
rating_1$KnowAxis = as.factor(rating_1$KnowAxis)
rating_1$X1990s2000s = as.factor(rating_1$X1990s2000s)
rating_1$CollegeMusic = as.factor(rating_1$CollegeMusic)
rating_1$APTheory = as.factor(rating_1$APTheory)
rating_1$Composing = as.factor(rating_1$Composing)
rating_1$PianoPlay = as.factor(rating_1$PianoPlay)
rating_1$GuitarPlay = as.factor(rating_1$GuitarPlay)

# Select fix factors for three main design factors
lm.c3.full = lm(Classical ~ Instrument * Harmony * Voice, data = rating_1)
stepAIC(lm.c3.full, k = 2)
lm.c3.final = lm(Classical ~ Harmony + Instrument + Voice + Instrument:Voice,
 data = rating_1)

# First from box plot then according to definitions and box plots
lm.cfix.trials = lm(Classical ~ Harmony + Instrument + Voice + Selfdeclare +
    OMSI + X16.minus.17 + ConsInstr + ConsNotes + PachListen +
    ClsListen + NoClass + Composing + PianoPlay +
    GuitarPlay + KnowAxis + KnowRob +X1990s2000s + X1990s2000s.minus.1960s197
0s, data = rating_1)

# Remove some fix effects
stepAIC(lm.cfix.trials, k=2)

# Final fix effect model
lm.cfix.final = lm(Classical ~ Harmony + Instrument + Voice + Selfdeclare + O
MSI +
    X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
    NoClass + Composing + PianoPlay + GuitarPlay + KnowAxis, data = rating_1)

# Fit lmer model
m.c.trials = lmer(Classical ~ Harmony + Instrument + Voice + Selfdeclare + OM
SI +
    X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
    NoClass + Composing + PianoPlay + GuitarPlay + KnowAxis+(1|Subject), data
 = rating_1, control = lmerControl(optimizer = 'bobyqa'),REML = FALSE)

# Select random effects and reselect fix effects
m.c.final = fitLMER.fnc(m.c.trials, ran.effects = c("(Instrument|Subject)", "
(Harmony|Subject)", "(Voice|Subject)", "(Selfdeclare|Subject)","(OMSI|Subjec
```

```
t)","(X16.minus.17|Subject)","(ConsInstr|Subject)", "(ConsNotes|Subject)", "
(PachListen|Subject)", "(ClsListen|Subject)", "(NoClass|Subject)", "(Composin
g|Subject)", "(PianoPlay|Subject)", "(GuitarPlay|Subject)","(KnowAxis|Subjec
t)"),  method = 'AIC')

# Model wil the interaction of KnowRob and Harmony, KnowAxis and Harmony
m.c.final.test1 = lmer(Classical ~ -1+ Harmony+ Instrument+KnowAxis:Harmony +
KnowRob:Harmony +Voice   + Selfdeclare + ConsInstr + PachListen + ClsListen +
 Composing + PianoPlay + GuitarPlay + KnowAxis  + KnowRob + (-1+ Harmony+Inst
rument | Subject), data = rating_1,control = lmerControl(optimizer = 'bobyqa
'),REML = FALSE )

# Graph the correlation between Harmony and Classical
rating_cor1 = rating_1%>% select(Harmony, Classical)
rating_cor1 <- fastDummies::dummy_cols(rating_cor1,select_columns = "Harmony
")

rating_cor1 = rating_cor1 %>%select(-Harmony)

cor(rating_cor1, method = "pearson")

##                    Classical Harmony_I-IV-V Harmony_I-V-IV Harmony_I-V-VI
## Classical       1.00000000    -0.04806789    -0.05085003      0.1319529
## Harmony_I-IV-V -0.04806789     1.00000000    -0.33449477     -0.3327511
## Harmony_I-V-IV -0.05085003    -0.33449477     1.00000000     -0.3327511
## Harmony_I-V-VI  0.13195294    -0.33275110    -0.33275110      1.0000000
## Harmony_IV-I-V -0.03271859    -0.33391355    -0.33391355     -0.3321729
##                Harmony_IV-I-V
## Classical         -0.03271859
## Harmony_I-IV-V    -0.33391355
## Harmony_I-V-IV    -0.33391355
## Harmony_I-V-VI    -0.33217290
## Harmony_IV-I-V     1.00000000

# Graph the correlation between Voice and Classical
rating_cor2 = rating_1 %>%select(Voice, Classical)
rating_cor2 = fastDummies::dummy_cols(rating_cor2, select_columns = "Voice")
rating_cor2 = rating_cor2 %>% select(-Voice)

cor(rating_cor2, method = "pearson")

##                  Classical Voice_contrary Voice_par3rd Voice_par5th
## Classical       1.00000000     0.06075679  -0.04123251  -0.01954448
## Voice_contrary  0.06075679     1.00000000  -0.49975508  -0.50048972
## Voice_par3rd   -0.04123251    -0.49975508   1.00000000  -0.49975508
## Voice_par5th   -0.01954448    -0.50048972  -0.49975508   1.00000000

# Compare the model with and without the interaction between KnowRob, KnowAxi
s and Harmony
anova(m.c.final,m.c.final.test1)
```

```
## Data: rating_1
## Models:
## m.c.final: Classical ~ Harmony + Instrument + Voice + Selfdeclare + X16.mi
nus.17 +
## m.c.final:     ConsInstr + ConsNotes + PachListen + ClsListen + NoClass +
## m.c.final:     Composing + PianoPlay + GuitarPlay + KnowAxis + (1 | Subjec
t) +
## m.c.final:     (Instrument | Subject) + (Harmony | Subject)
## m.c.final.test1: Classical ~ -1 + Harmony + Instrument + KnowAxis:Harmony
+ KnowRob:Harmony +
## m.c.final.test1:     Voice + Selfdeclare + ConsInstr + PachListen + ClsLis
ten +
## m.c.final.test1:     Composing + PianoPlay + GuitarPlay + KnowAxis + KnowR
ob +
## m.c.final.test1:     (-1 + Harmony + Instrument | Subject)
##                 Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## m.c.final       67 6112.2 6469.6 -2989.1   5978.2
## m.c.final.test1 79 6093.0 6514.4 -2967.5   5935.0 43.29     12  2.017e-05
##
## m.c.final
## m.c.final.test1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 1

```
# The summary for the better model
summary(m.c.final.test1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ -1 + Harmony + Instrument + KnowAxis:Harmony + KnowRob:Harmony
 +
##     Voice + Selfdeclare + ConsInstr + PachListen + ClsListen +
##     Composing + PianoPlay + GuitarPlay + KnowAxis + KnowRob +
##     (-1 + Harmony + Instrument | Subject)
##    Data: rating_1
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   6093.0   6514.4  -2967.5   5935.0     1453
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6345 -0.5681  0.0152  0.5645  3.8129
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  HarmonyI-IV-V    0.4179   0.6465
##           HarmonyI-V-IV    0.7587   0.8711    0.97
```

```
##           HarmonyI-V-VI    1.7053   1.3059    0.51  0.69
##           HarmonyIV-I-V    0.4149   0.6441    0.92  0.91  0.60
##           Instrumentpiano  2.0539   1.4331   -0.42 -0.53 -0.55 -0.47
##           Instrumentstring 3.5586   1.8864   -0.80 -0.92 -0.92 -0.85  0.61
##  Residual                  2.4007   1.5494
## Number of obs: 1532, groups:  Subject, 43
##
## Fixed effects:
##                          Estimate Std. Error t value
## HarmonyI-IV-V           2.3956198  0.8728693   2.745
## HarmonyI-V-IV           2.4477037  0.8770508   2.791
## HarmonyI-V-VI           3.1666908  0.8911739   3.553
## HarmonyIV-I-V           2.5038376  0.8725928   2.869
## Instrumentpiano         1.6007023  0.2393616   6.687
## Instrumentstring        3.5271238  0.3036267  11.617
## Voicepar3rd            -0.3856677  0.0970333  -3.975
## Voicepar5th            -0.3027265  0.0969731  -3.122
## Selfdeclare2           -1.1280789  0.4357443  -2.589
## Selfdeclare3           -1.7979206  0.4775014  -3.765
## Selfdeclare4           -0.2354891  0.6193972  -0.380
## Selfdeclare5           -6.8355382  0.9674489  -7.066
## Selfdeclare6           -2.3868523  1.2161393  -1.963
## ConsInstr0.67           1.9858818  0.9075098   2.188
## ConsInstr1             -1.8216575  0.5783951  -3.150
## ConsInstr1.67           0.0618574  0.4766702   0.130
## ConsInstr2.33           0.0568428  0.4083447   0.139
## ConsInstr2.67           1.5470107  0.4576518   3.380
## ConsInstr3             -0.6093216  0.5176882  -1.177
## ConsInstr3.33          -1.9941687  0.6808136  -2.929
## ConsInstr3.67          -0.3228766  0.4843840  -0.667
## ConsInstr4              1.4812703  0.6334952   2.338
## ConsInstr4.33           0.5047916  0.4348765   1.161
## ConsInstr5              0.3140765  0.4163215   0.754
## PachListen3             0.7305734  0.4808619   1.519
## PachListen5             1.9183677  0.3928462   4.883
## ClsListen1              0.1019741  0.3308936   0.308
## ClsListen3              0.7018131  0.3790424   1.852
## ClsListen4              3.1058559  0.8415697   3.691
## ClsListen5             -1.9267906  0.5310089  -3.629
## Composing1              0.7839123  0.3148308   2.490
## Composing2             -0.0005438  0.5263683  -0.001
## Composing3             -2.6275395  0.5622118  -4.674
## Composing4              2.0127585  0.3845061   5.235
## PianoPlay1              0.8332166  0.2789695   2.987
## PianoPlay4              3.5148026  0.5367585   6.548
## PianoPlay5              1.2863744  0.5480561   2.347
## GuitarPlay1             4.4623355  0.7162266   6.230
## GuitarPlay2            -2.5456493  0.9443145  -2.696
## GuitarPlay4             2.3398697  1.0192611   2.296
## GuitarPlay5             0.0586711  0.4923465   0.119
```

```
## KnowAxis1               -2.3364110  1.7260995  -1.354
## KnowAxis5                0.1834731  0.3559355   0.515
## KnowRob1                 2.1685321  0.6835156   3.173
## KnowRob5                -0.5357760  0.4497033  -1.191
## HarmonyI-V-IV:KnowAxis1  1.4470439  0.8249082   1.754
## HarmonyI-V-VI:KnowAxis1 -1.4875908  1.3104638  -1.135
## HarmonyIV-I-V:KnowAxis1  0.6147052  0.8633453   0.712
## HarmonyI-V-IV:KnowAxis5 -0.1538648  0.2907568  -0.529
## HarmonyI-V-VI:KnowAxis5 -1.1018325  0.4616157  -2.387
## HarmonyIV-I-V:KnowAxis5 -0.3489073  0.3040655  -1.147
## HarmonyI-V-IV:KnowRob1   0.0404009  0.4047803   0.100
## HarmonyI-V-VI:KnowRob1   0.9296492  0.6473257   1.436
## HarmonyIV-I-V:KnowRob1   0.2286105  0.4242990   0.539
## HarmonyI-V-IV:KnowRob5  -0.3986574  0.3670341  -1.086
## HarmonyI-V-VI:KnowRob5   1.9027494  0.5786195   3.288
## HarmonyIV-I-V:KnowRob5   0.1014195  0.3818912   0.266
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 2 columns / coeffi
cients
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Table 2

Residuals for classical rating model

```
#############################################################
#
# Residuals!


attach(rating_1)
source("residual-functions.r")

resid.marg <- r.marg(m.c.final.test1)
resid.cond <- r.cond(m.c.final.test1)
resid.reff <- r.reff(m.c.final.test1)

par(mfrow=c(2,2))
plot(resid.marg,xlab="Index",ylab="Marginal Residuals")
abline(0,0)
plot(resid.cond,xlab="Index",ylab="Conditional Residuals")
abline(0,0)
plot(resid.reff,xlab="Index",ylab="Random Effects")
abline(0,0)

sch <- as.numeric(Subject)
index <- sch
for (j in unique(sch)) {
  len <- sum(sch==j)
  index[sch==j] <- 1:len
```

```
}

new.data <- data.frame(index,resid.marg,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)

new.data <- data.frame(index,resid.cond,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)

new.data <- data.frame(index,resid.reff,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)

detach(rating_1)

#######
```

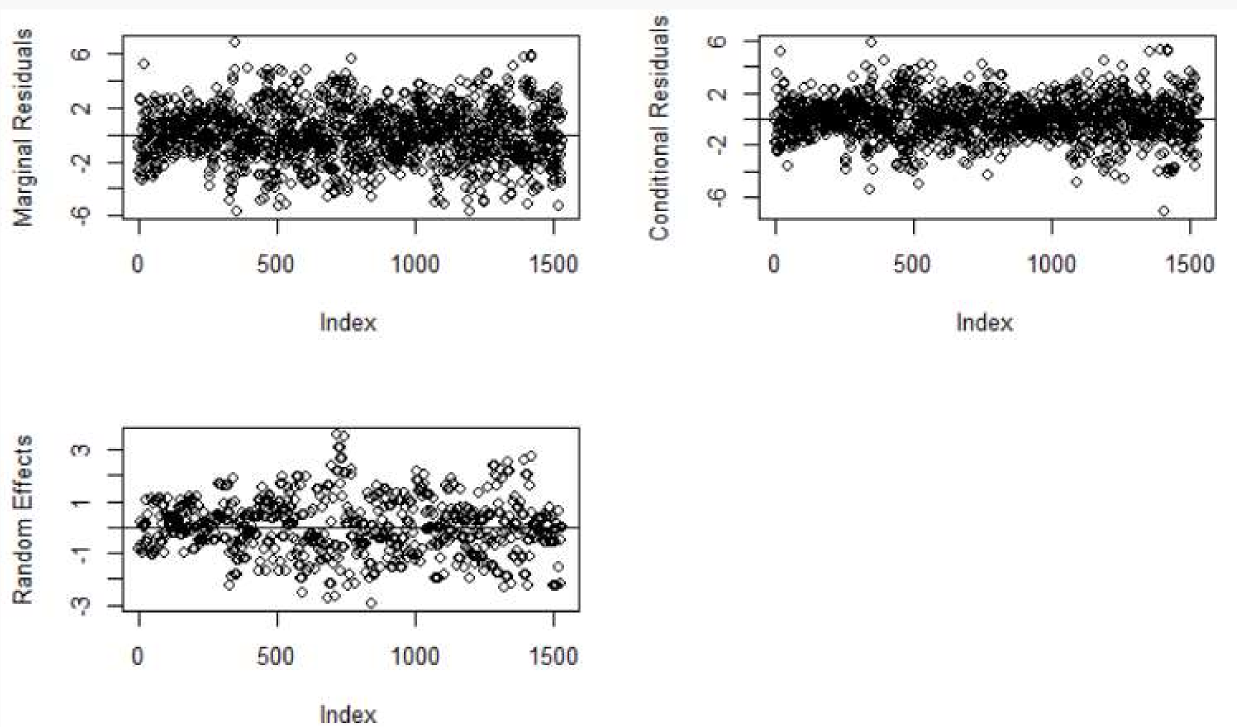


Figure 6

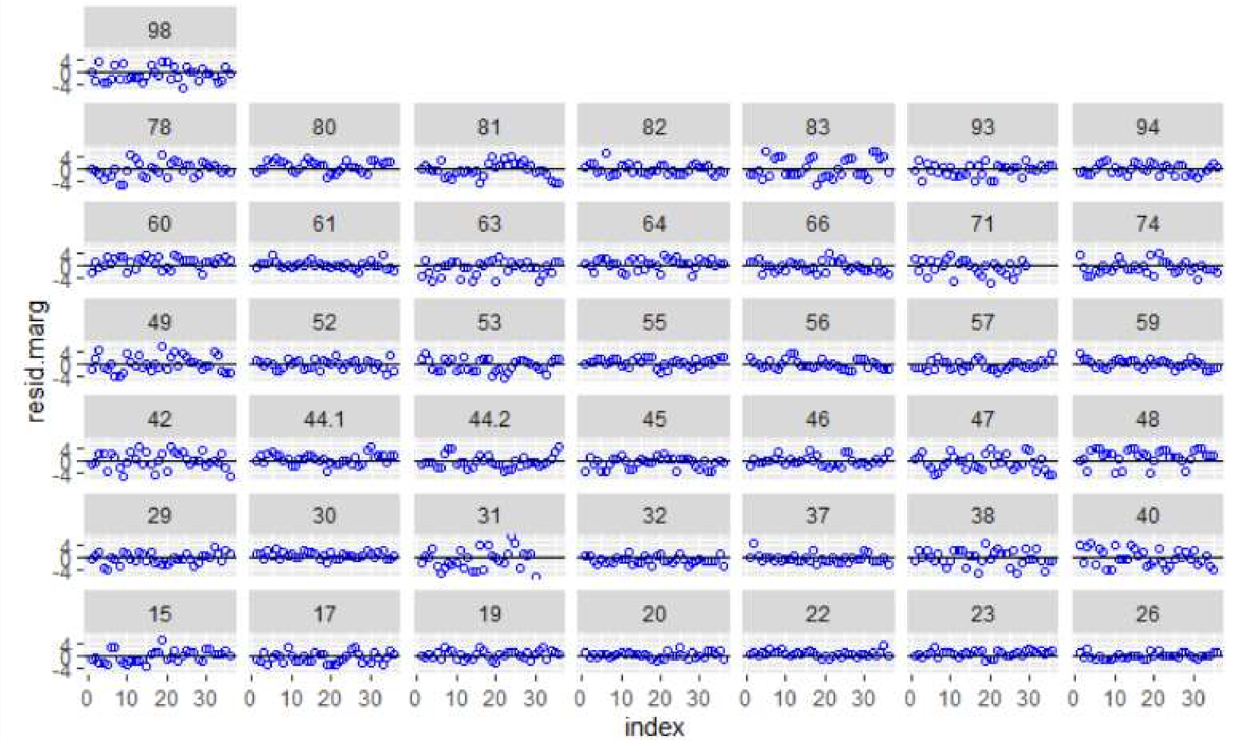


Figure 7

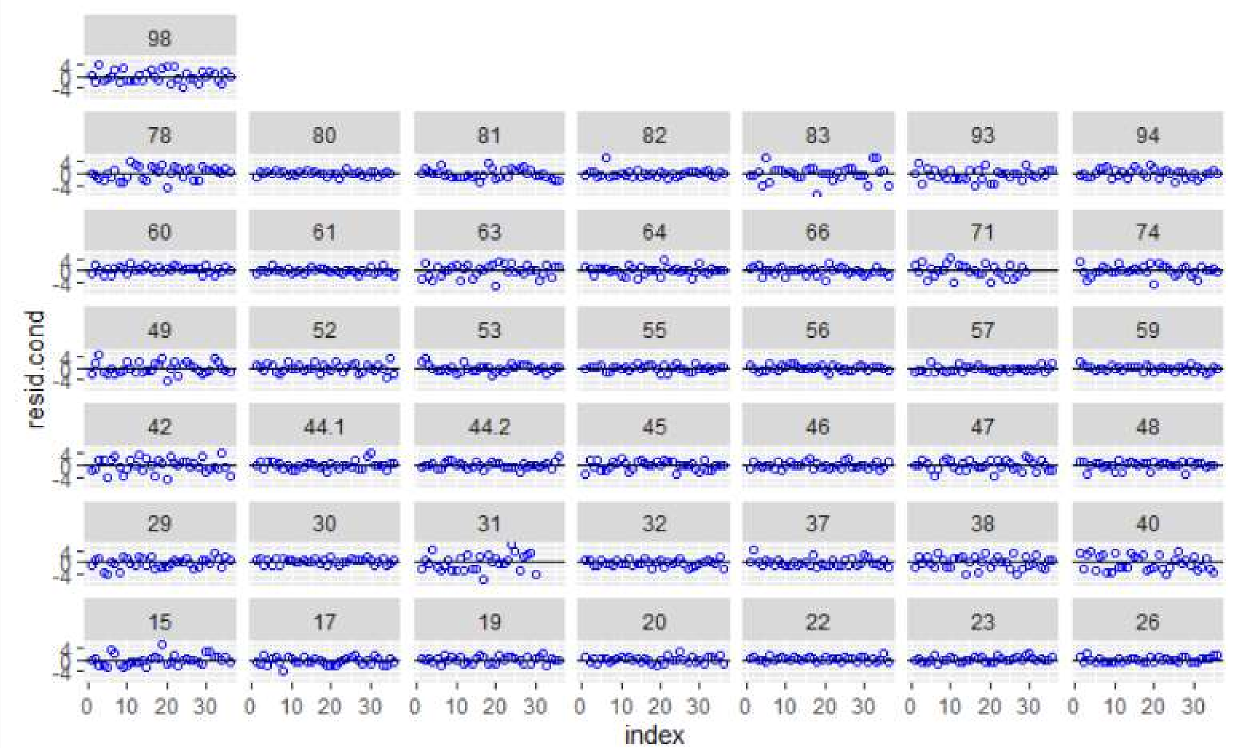


Figure 8

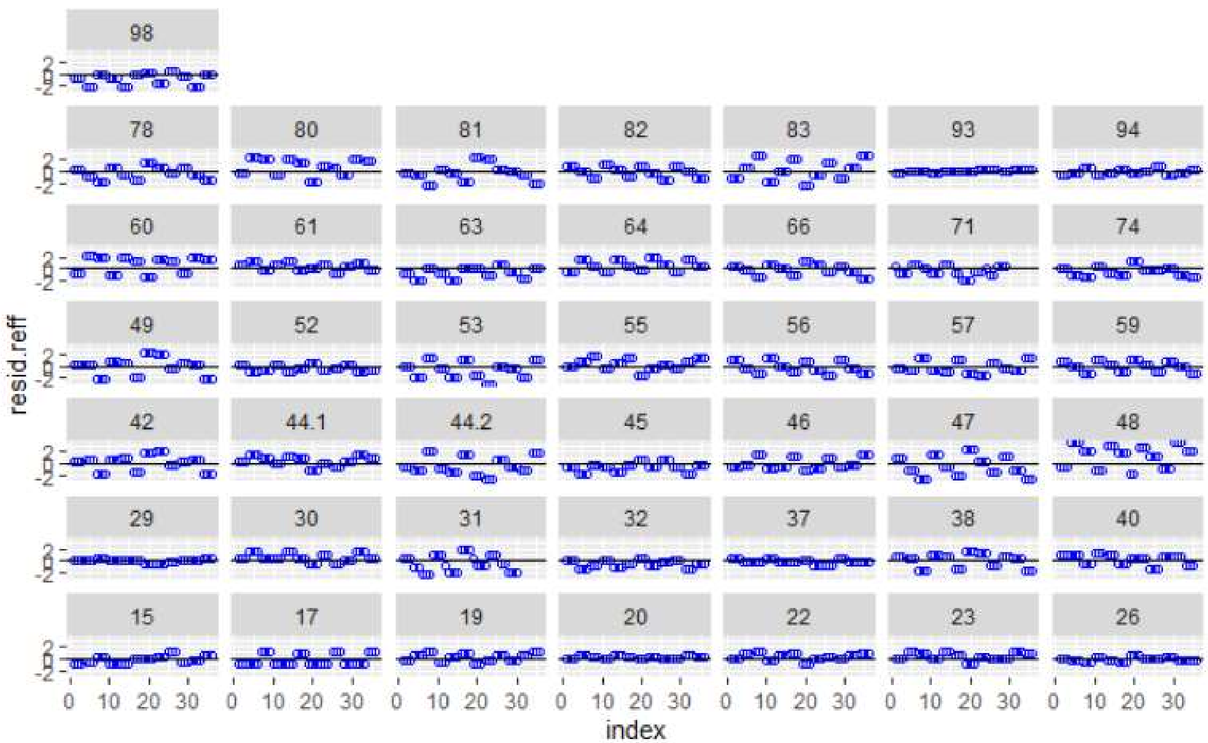


Figure 9

Classical rating with musicians

```
# dichotomize self declare at 2
rating_41b = rating_1
rating_41b$Selfdeclare = as.numeric(rating_41b$Selfdeclare)
rating_41b = rating_41b %>% mutate(Musician = case_when(Selfdeclare>2 ~ "True
", Selfdeclare<=2 ~ "False" ))
rating_41b$Musician = as.factor(rating_41b$Musician)

# find the best model
lm.cwith.musician = lm(Classical ~ (KnowAxis:Harmony +KnowRob:Harmony +Voice
+ Harmony + Instrument + ConsInstr + PachListen + ClsListen + Composing + Pia
noPlay + GuitarPlay + KnowAxis  + KnowRob)*Musician , data = rating_41b)

stepAIC(lm.cwith.musician, k = 2)

m.cwith.musician = lmer(Classical ~ Voice + Harmony + Instrument + ConsInstr
+ ClsListen +
    Composing + PianoPlay + KnowAxis + KnowRob + Musician + Harmony:KnowAxis
+
    Harmony:KnowRob + Harmony:Musician + Instrument:Musician +
    ConsInstr:Musician + ClsListen:Musician + KnowAxis:Musician +
    KnowRob:Musician + Harmony:KnowAxis:Musician+(Instrument + Harmony|Subjec
t ), data = rating_41b,  REML = FALSE, control = lmerControl(optimizer = 'bob
yqa'))
```

```
m.cwith.musician.test1 = update(m.cwith.musician, .~. +(1|Musician) )

m.cwith.musician.test2 = update(m.cwith.musician, .~.   + (Musician | Subjec
t) )

m.cwith.musician.test3 = update(m.cwith.musician.test2, .~. + (Musician | Sub
ject) )

anova(m.cwith.musician.test1, m.cwith.musician)

## Data: rating_41b
## Models:
## m.cwith.musician: Classical ~ Voice + Harmony + Instrument + ConsInstr + C
lsListen +
## m.cwith.musician:     Composing + PianoPlay + KnowAxis + KnowRob + Musicia
n + Harmony:KnowAxis +
## m.cwith.musician:     Harmony:KnowRob + Harmony:Musician + Instrument:Musi
cian +
## m.cwith.musician:     ConsInstr:Musician + ClsListen:Musician + KnowAxis:M
usician +
## m.cwith.musician:     KnowRob:Musician + Harmony:KnowAxis:Musician + (Inst
rument +
## m.cwith.musician:     Harmony | Subject)
## m.cwith.musician.test1: Classical ~ Voice + Harmony + Instrument + ConsIns
tr + ClsListen +
## m.cwith.musician.test1:     Composing + PianoPlay + KnowAxis + KnowRob + M
usician + (Instrument +
## m.cwith.musician.test1:     Harmony | Subject) + (1 | Musician) + Harmony:
KnowAxis +
## m.cwith.musician.test1:     Harmony:KnowRob + Harmony:Musician + Instrumen
t:Musician +
## m.cwith.musician.test1:     ConsInstr:Musician + ClsListen:Musician + Know
Axis:Musician +
## m.cwith.musician.test1:     KnowRob:Musician + Harmony:KnowAxis:Musician
##                         Df    AIC    BIC  logLik deviance  Chisq Chi Df
## m.cwith.musician        89 6062.5 6537.3 -2942.3   5884.5
## m.cwith.musician.test1 90 6064.4 6544.4 -2942.2   5884.4 0.1776      1
##                         Pr(>Chisq)
## m.cwith.musician
## m.cwith.musician.test1     0.6735
```

Table 3

```
anova(m.cwith.musician.test2, m.cwith.musician)

## Data: rating_41b
## Models:
## m.cwith.musician: Classical ~ Voice + Harmony + Instrument + ConsInstr + C
lsListen +
## m.cwith.musician:     Composing + PianoPlay + KnowAxis + KnowRob + Musicia
```

```
n + Harmony:KnowAxis +
## m.cwith.musician:     Harmony:KnowRob + Harmony:Musician + Instrument:Musi
cian +
## m.cwith.musician:     ConsInstr:Musician + ClsListen:Musician + KnowAxis:M
usician +
## m.cwith.musician:     KnowRob:Musician + Harmony:KnowAxis:Musician + (Inst
rument +
## m.cwith.musician:     Harmony | Subject)
## m.cwith.musician.test2: Classical ~ Voice + Harmony + Instrument + ConsIns
tr + ClsListen +
## m.cwith.musician.test2:     Composing + PianoPlay + KnowAxis + KnowRob + M
usician + (Instrument +
## m.cwith.musician.test2:     Harmony | Subject) + (Musician | Subject) + Ha
rmony:KnowAxis +
## m.cwith.musician.test2:     Harmony:KnowRob + Harmony:Musician + Instrumen
t:Musician +
## m.cwith.musician.test2:     ConsInstr:Musician + ClsListen:Musician + Know
Axis:Musician +
## m.cwith.musician.test2:     KnowRob:Musician + Harmony:KnowAxis:Musician
##                         Df    AIC    BIC  logLik deviance Chisq Chi Df
## m.cwith.musician        89 6062.5 6537.3 -2942.3   5884.5
## m.cwith.musician.test2 92 6098.1 6588.8 -2957.0   5914.1     0      3
##                         Pr(>Chisq)
## m.cwith.musician
## m.cwith.musician.test2          1
```

Table 4

```
anova(m.cwith.musician.test3, m.cwith.musician.test2)
```

```
## Data: rating_41b
## Models:
## m.cwith.musician.test3: Classical ~ Voice + Harmony + Instrument + ConsIns
tr + ClsListen +
## m.cwith.musician.test3:     Composing + PianoPlay + KnowAxis + KnowRob + M
usician + (Instrument +
## m.cwith.musician.test3:     Harmony | Subject) + (Musician | Subject) + Ha
rmony:KnowAxis +
## m.cwith.musician.test3:     Harmony:KnowRob + Harmony:Musician + Instrumen
t:Musician +
## m.cwith.musician.test3:     ConsInstr:Musician + ClsListen:Musician + Know
Axis:Musician +
## m.cwith.musician.test3:     KnowRob:Musician + Harmony:KnowAxis:Musician
## m.cwith.musician.test2: Classical ~ Voice + Harmony + Instrument + ConsIns
tr + ClsListen +
## m.cwith.musician.test2:     Composing + PianoPlay + KnowAxis + KnowRob + M
usician + (Instrument +
## m.cwith.musician.test2:     Harmony | Subject) + (Musician | Subject) + Ha
rmony:KnowAxis +
## m.cwith.musician.test2:     Harmony:KnowRob + Harmony:Musician + Instrumen
```

```
t:Musician +
## m.cwith.musician.test2:     ConsInstr:Musician + ClsListen:Musician + Know
Axis:Musician +
## m.cwith.musician.test2:     KnowRob:Musician + Harmony:KnowAxis:Musician
##                          Df    AIC    BIC logLik deviance Chisq Chi Df
## m.cwith.musician.test3 92 6098.1 6588.8  -2957   5914.1
## m.cwith.musician.test2 92 6098.1 6588.8  -2957   5914.1     0      0
##                          Pr(>Chisq)
## m.cwith.musician.test3
## m.cwith.musician.test2          1
```

Table 5

```
summary(m.cwith.musician)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ Voice + Harmony + Instrument + ConsInstr + ClsListen +
##     Composing + PianoPlay + KnowAxis + KnowRob + Musician + Harmony:KnowAx
is +
##     Harmony:KnowRob + Harmony:Musician + Instrument:Musician +
##     ConsInstr:Musician + ClsListen:Musician + KnowAxis:Musician +
##     KnowRob:Musician + Harmony:KnowAxis:Musician + (Instrument +
##     Harmony | Subject)
##    Data: rating_41b
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   6062.5   6537.3  -2942.3   5884.5     1443
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6379 -0.5756  0.0068  0.5631  3.8614
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      0.64978  0.8061
##           Instrumentpiano  1.91083  1.3823   -0.81
##           Instrumentstring 3.27158  1.8088   -0.90  0.59
##           HarmonyI-V-IV    0.09897  0.3146    0.86 -0.74 -0.94
##           HarmonyI-V-VI    0.90515  0.9514    0.36 -0.40 -0.60  0.77
##           HarmonyIV-I-V    0.04267  0.2066    0.14 -0.34 -0.15  0.31  0.25
##  Residual                  2.37692  1.5417
## Number of obs: 1532, groups:  Subject, 43
##
## Fixed effects:
##                                       Estimate Std. Error t value
## (Intercept)                          3.208e+00  5.810e-01   5.521
## Voicepar3rd                         -3.855e-01  9.655e-02  -3.993
## Voicepar5th                         -3.025e-01  9.649e-02  -3.135
```

```
## HarmonyI-V-IV                         5.023e-04  1.863e-01   0.003
## HarmonyI-V-VI                         1.366e-01  2.789e-01   0.490
## HarmonyIV-I-V                        -2.820e-02  1.809e-01  -0.156
## Instrumentpiano                       1.868e+00  3.174e-01   5.885
## Instrumentstring                      4.030e+00  3.999e-01  10.078
## ConsInstr0.67                         2.970e+00  7.049e-01   4.214
## ConsInstr1                            7.569e-01  4.942e-01   1.532
## ConsInstr1.67                        -4.726e-01  4.414e-01  -1.071
## ConsInstr2.33                         7.900e-01  3.590e-01   2.201
## ConsInstr2.67                         4.475e-01  3.661e-01   1.222
## ConsInstr3                           -5.898e-01  5.099e-01  -1.157
## ConsInstr3.33                        -4.250e+00  7.748e-01  -5.486
## ConsInstr3.67                         1.282e+00  5.099e-01   2.515
## ConsInstr4                            1.180e+00  6.126e-01   1.925
## ConsInstr4.33                        -2.383e+00  4.457e-01  -5.347
## ConsInstr5                            1.272e+00  4.390e-01   2.898
## ClsListen1                           -1.323e+00  3.153e-01  -4.196
## ClsListen3                            1.148e+00  4.704e-01   2.440
## ClsListen4                           -2.177e+01  2.672e+00  -8.147
## ClsListen5                           -3.206e+00  4.902e-01  -6.540
## Composing1                            5.359e-01  2.465e-01   2.174
## Composing2                            6.648e+00  6.042e-01  11.003
## Composing3                           -1.708e+01  1.911e+00  -8.941
## Composing4                            2.160e+00  3.345e-01   6.457
## Composing5                           -5.238e+00  7.378e-01  -7.098
## PianoPlay1                           -1.464e+00  3.416e-01  -4.286
## PianoPlay4                            6.157e+00  4.307e-01  14.296
## PianoPlay5                            6.430e+00  6.989e-01   9.200
## KnowAxis1                             6.383e+00  1.001e+00   6.379
## KnowAxis5                             2.112e+00  4.101e-01   5.148
## KnowRob1                              1.705e+01  2.091e+00   8.158
## KnowRob5                              3.190e-01  5.287e-01   0.603
## MusicianTrue                          2.454e+00  8.701e-01   2.820
## HarmonyI-V-IV:KnowAxis1               1.410e+00  8.286e-01   1.701
## HarmonyI-V-VI:KnowAxis1              -1.743e+00  1.195e+00  -1.458
## HarmonyIV-I-V:KnowAxis1               5.067e-01  8.435e-01   0.601
## HarmonyI-V-IV:KnowAxis5               4.669e-02  4.012e-01   0.116
## HarmonyI-V-VI:KnowAxis5              -2.698e-01  5.794e-01  -0.466
## HarmonyIV-I-V:KnowAxis5               1.712e-01  4.090e-01   0.419
## HarmonyI-V-IV:KnowRob1                2.216e-01  4.234e-01   0.523
## HarmonyI-V-VI:KnowRob1                8.768e-01  6.158e-01   1.424
## HarmonyIV-I-V:KnowRob1                4.246e-01  4.318e-01   0.984
## HarmonyI-V-IV:KnowRob5               -3.770e-01  3.741e-01  -1.008
## HarmonyI-V-VI:KnowRob5                1.618e+00  5.370e-01   3.014
## HarmonyIV-I-V:KnowRob5                3.071e-02  3.791e-01   0.081
## HarmonyI-V-IV:MusicianTrue            7.649e-02  2.896e-01   0.264
## HarmonyI-V-VI:MusicianTrue            1.501e+00  4.320e-01   3.475
## HarmonyIV-I-V:MusicianTrue            2.979e-01  2.823e-01   1.055
## Instrumentpiano:MusicianTrue         -5.723e-01  4.653e-01  -1.230
## Instrumentstring:MusicianTrue        -1.080e+00  5.859e-01  -1.843
```

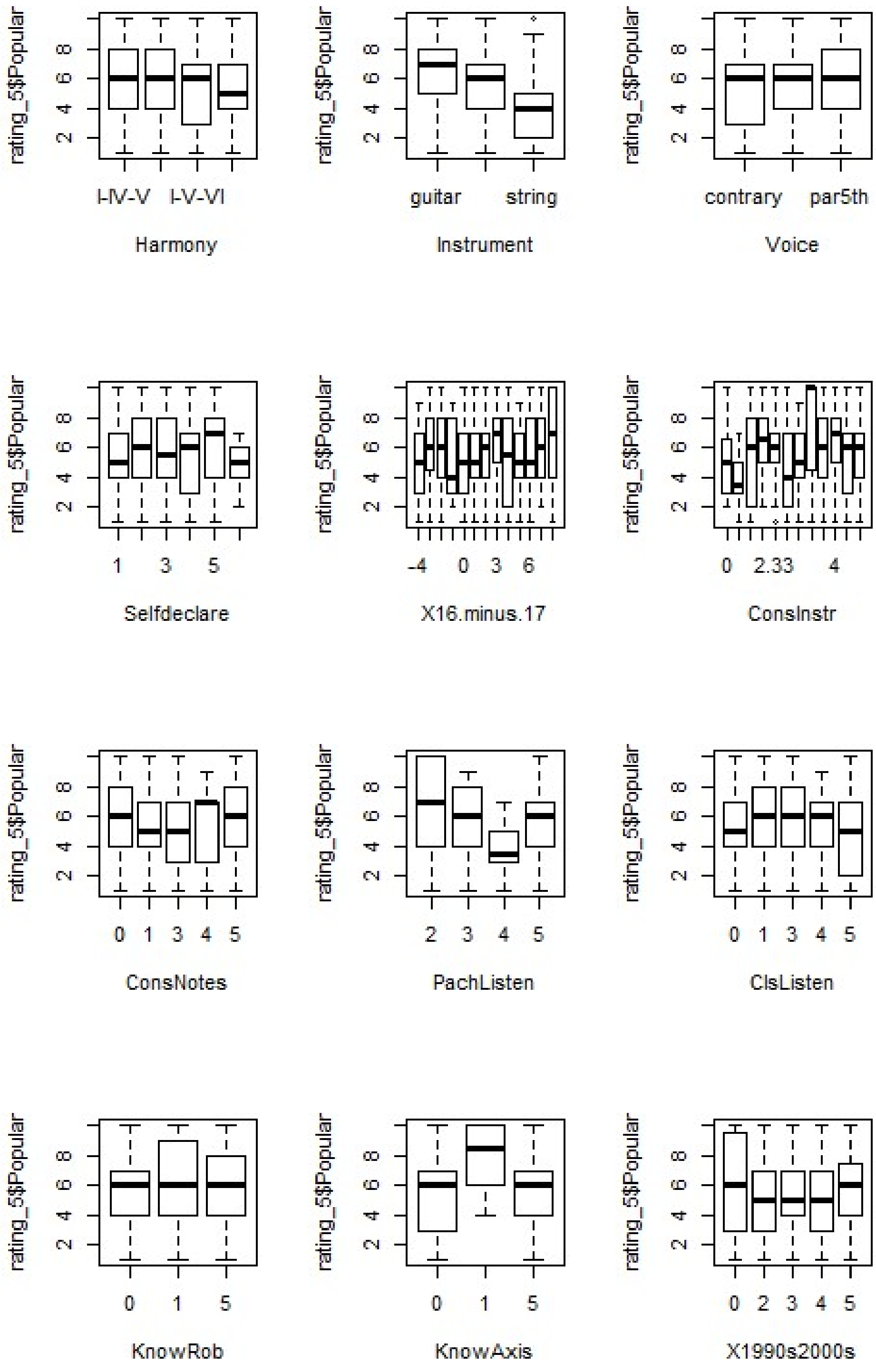
```
## ConsInstr1:MusicianTrue                  1.085e+01  1.497e+00   7.248
## ConsInstr1.67:MusicianTrue               1.314e+01  1.727e+00   7.607
## ConsInstr3:MusicianTrue                 -2.354e+00  6.720e-01  -3.503
## ConsInstr3.67:MusicianTrue               5.746e-01  5.254e-01   1.094
## ConsInstr4.33:MusicianTrue               2.168e+01  2.478e+00   8.751
## ClsListen1:MusicianTrue                 -1.998e+00  8.419e-01  -2.373
## ClsListen3:MusicianTrue                 -5.085e+00  9.096e-01  -5.590
## ClsListen5:MusicianTrue                 -1.693e+00  1.564e+00  -1.082
## KnowAxis5:MusicianTrue                   4.266e+00  7.882e-01   5.412
## KnowRob1:MusicianTrue                   -2.447e+01  2.651e+00  -9.230
## KnowRob5:MusicianTrue                   -1.583e+01  1.760e+00  -8.991
## HarmonyI-V-IV:KnowAxis5:MusicianTrue -3.880e-01  5.401e-01  -0.718
## HarmonyI-V-VI:KnowAxis5:MusicianTrue -1.579e+00  7.806e-01  -2.023
## HarmonyIV-I-V:KnowAxis5:MusicianTrue -9.954e-01  5.510e-01  -1.807
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 11 columns / coeff
icients
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Table 6

Popular model

Data preprocessing

```
# Remove x1instr and x2instr and Classical
rating_5 = rating[c(1:22,26)]
# Log transform OMSI
rating_5$OMSI = log(rating_5$OMSI)

# I drop Instr.minus.Notes, because the distribution is not normal and it is
highly correlated with ConsNote
rating_5= rating_5[c(1:9,11:23)]

# I also drop the Classicle rate == 0, NA and 19
rating_5 = rating_5 %>% filter(rating_5$Popular !=0, !is.na(rating_5$Popula
r), rating_5$Popular != 3.5, rating_5$Popular != 4.2, rating_5$Popular != 4.
6, rating_5$Popular != 6.8)

rating_5 = rating_5 %>% mutate(Popular=replace(Popular, Popular==19, 10))
rating_5 = na.omit(rating_5)

# boxplot
par(mfrow = c(2,3))
for (i in names(rating_5%>%dplyr::select(-OMSI,-Popular, - Subject))) {boxplo
t(rating_5$Popular ~ rating_5[,i], xlab = i)}
```


rating_5$Popular
Harmony
I-IV-V
I-V-VI
Instrument
guitar
string
Voice
contrary
par5th
Selfdeclare
X16.minus.17
ConsInstr
2.33
ConsNotes
PachListen
ClsListen
KnowRob
KnowAxis
X1990s2000s

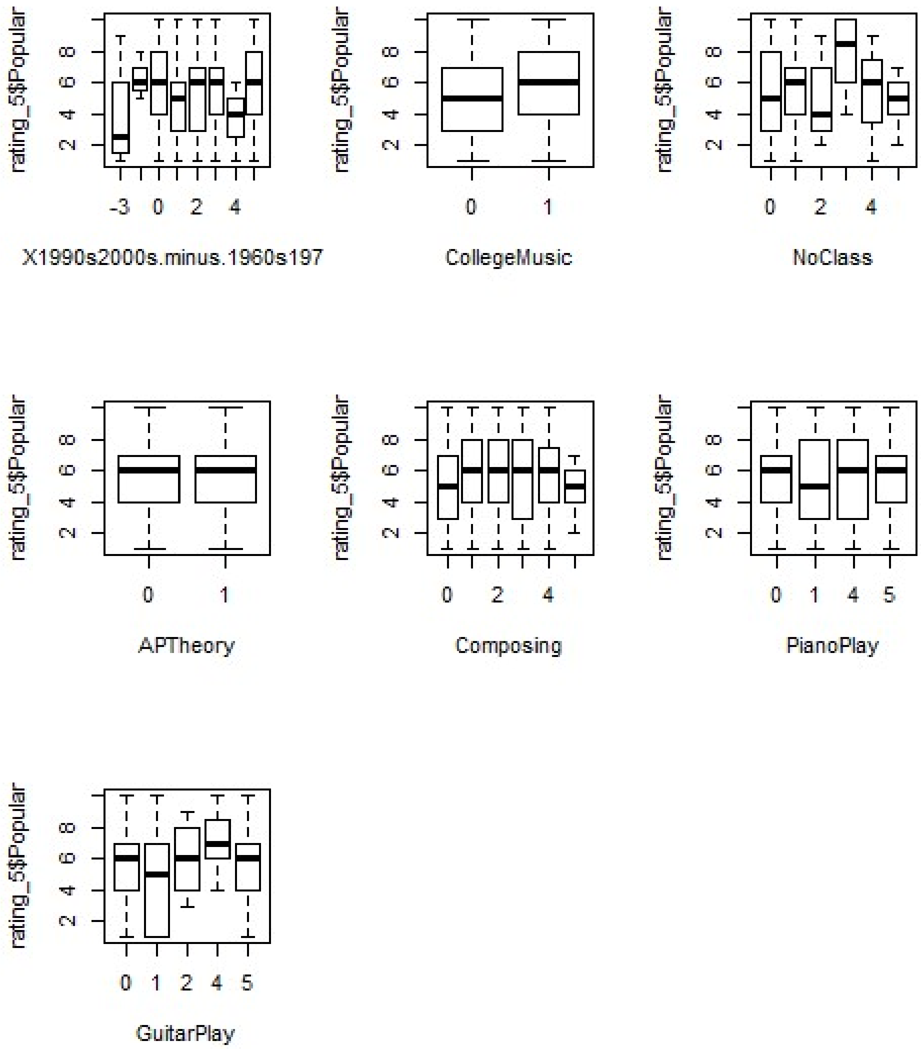

rating_5$Popular
X1990s2000s.minus.1960s197
CollegeMusic
NoClass
APTheory
Composing
PianoPlay
GuitarPlay

Figure 10

Popular rating final model

```
# select fixed effects for three main factors
lm.51 = lm(Popular ~ Instrument * Harmony * Voice, data = rating_5)
stepAIC(lm.51, k = 2)
# final fixed effects for three main factors
lm.52 = lm(Popular ~ Harmony + Instrument, data = rating_5)

# transform the variables into factors
rating_5$Selfdeclare = as.factor(rating_5$Selfdeclare)
rating_5$ConsInstr = as.factor(rating_5$ConsInstr)
rating_5$ConsNotes = as.factor(rating_5$ConsNotes)
rating_5$PachListen = as.factor(rating_5$PachListen)
rating_5$ClsListen = as.factor(rating_5$ClsListen)
rating_5$KnowRob = as.factor(rating_5$KnowRob)
rating_5$KnowAxis = as.factor(rating_5$KnowAxis)
rating_5$X1990s2000s = as.factor(rating_5$X1990s2000s)
rating_5$CollegeMusic = as.factor(rating_5$CollegeMusic)
rating_5$APTheory = as.factor(rating_5$APTheory)
rating_5$Composing = as.factor(rating_5$Composing)
rating_5$PianoPlay = as.factor(rating_5$PianoPlay)
rating_5$GuitarPlay = as.factor(rating_5$GuitarPlay)

# Start looking for fix effect
lm.53 = lm(Popular ~ ConsInstr + PachListen + KnowAxis +X1990s2000s.minus.196
0s1970s + NoClass +GuitarPlay + Voice +Instrument + Selfdeclare+ X16.minus.17
 + ConsNotes+(KnowAxis + KnowRob)*Harmony, data = rating_5 )

stepAIC(lm.53, k=log(1517), scope=list(lower=Popular ~Voice + Instrument + Ha
rmony, upper=lm.53))

# start random effect
m.53 = lmer(Popular ~ ConsInstr + PachListen + KnowAxis + GuitarPlay + Voice
+
    Instrument + ConsNotes + KnowRob + Harmony+Voice +(1|Subject), data = rat
ing_5, REML = FALSE)

# Feature selection for both fix and random

m.54 = fitLMER.fnc(m.53, ran.effects = c("(Instrument|Subject)","((Harmony|Su
bject))","(PachListen|Subject)","(ConsInstr|Subject)","(Voice|Subject)", "(Kn
owRob|Subject)", "(KnowAxis|Subject)", "(ConsNotes|Subject)","(NoClass|Subjec
t)", "(GuitarPlay|Subject)"), method = 'BIC', keep.single.factors = TRUE)

m.final.Popular = lmer(Popular ~ -1 + Instrument+ Voice+ Harmony +ConsInstr +
 PachListen + KnowAxis + GuitarPlay   + ConsNotes + KnowRob + (Harmony +Instr
ument | Subject), data = rating_5,REML = FALSE, control = lmerControl(optimiz
er = 'bobyqa'))

summary(m.final.Popular)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Popular ~ -1 + Instrument + Voice + Harmony + ConsInstr + PachListen +
##     KnowAxis + GuitarPlay + ConsNotes + KnowRob + (Harmony +
##     Instrument | Subject)
##    Data: rating_5
## Control: lmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   6110.4   6403.3  -3000.2   6000.4     1462
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7696 -0.5922  0.0161  0.5905  3.3589
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Subject  (Intercept)      0.5746   0.7581
##           HarmonyI-V-IV    0.1729   0.4158    0.56
##           HarmonyI-V-VI    0.7682   0.8765    0.05 -0.03
##           HarmonyIV-I-V    0.3951   0.6286   -0.15 -0.50 -0.21
##           Instrumentpiano  1.7505   1.3231   -0.05 -0.38 -0.30 -0.38
##           Instrumentstring 2.5725   1.6039   -0.68 -0.46 -0.17 -0.40  0.72
##  Residual                  2.5423   1.5945
## Number of obs: 1517, groups:  Subject, 43
##
## Fixed effects:
##                  Estimate Std. Error t value
## Instrumentguitar  7.92085    0.72401  10.940
## Instrumentpiano   6.76521    0.74991   9.021
## Instrumentstring  4.95368    0.73875   6.705
## Voicepar3rd       0.16291    0.10041   1.622
## Voicepar5th       0.21325    0.10042   2.124
## HarmonyI-V-IV     0.01052    0.13223   0.080
## HarmonyI-V-VI    -0.28360    0.17700  -1.602
## HarmonyIV-I-V    -0.28512    0.15016  -1.899
## ConsInstr0.67    -3.80847    0.96913  -3.930
## ConsInstr1        1.53008    0.62020   2.467
## ConsInstr1.67     1.32121    0.69004   1.915
## ConsInstr2.33     0.76813    0.58425   1.315
## ConsInstr2.67    -0.86561    0.69813  -1.240
## ConsInstr3       -0.67123    0.65439  -1.026
## ConsInstr3.33     2.23771    0.82223   2.722
## ConsInstr3.67     0.06521    0.73337   0.089
## ConsInstr4        0.39788    0.79263   0.502
## ConsInstr4.33     0.15072    0.65787   0.229
## ConsInstr5        0.64820    0.67732   0.957
## PachListen3      -0.02093    0.66379  -0.032
## PachListen5      -1.45725    0.50934  -2.861
## KnowAxis1        -0.19880    0.84963  -0.234
```

```
## KnowAxis5         0.56688    0.32975   1.719
## GuitarPlay1      -1.00193    0.59539  -1.683
## GuitarPlay2       1.32173    0.59764   2.212
## GuitarPlay4       0.79992    0.61092   1.309
## GuitarPlay5      -0.02243    0.31982  -0.070
## ConsNotes1       -0.36001    0.47352  -0.760
## ConsNotes3       -0.42590    0.39630  -1.075
## ConsNotes4        0.10371    0.65439   0.158
## ConsNotes5        0.12205    0.45267   0.270
## KnowRob1          0.61434    0.35879   1.712
## KnowRob5         -0.59852    0.33891  -1.766
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coeffic
ient
## convergence code: 0
## boundary (singular) fit: see ?isSingular

                                   Table 7
```

Popular rating final model residuals

```
##############################################################
#
# Residuals!


attach(rating_5)
source("residual-functions.r")

resid.marg <- r.marg(m.final.Popular)
resid.cond <- r.cond(m.final.Popular)
resid.reff <- r.reff(m.final.Popular)

par(mfrow=c(2,2))
plot(resid.marg,xlab="Index",ylab="Marginal Residuals")
abline(0,0)
plot(resid.cond,xlab="Index",ylab="Conditional Residuals")
abline(0,0)
plot(resid.reff,xlab="Index",ylab="Random Effects")
abline(0,0)

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

new.data <- data.frame(index,resid.marg,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)

new.data <- data.frame(index,resid.cond,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)

new.data <- data.frame(index,resid.reff,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)

detach(rating_5)

#######
```

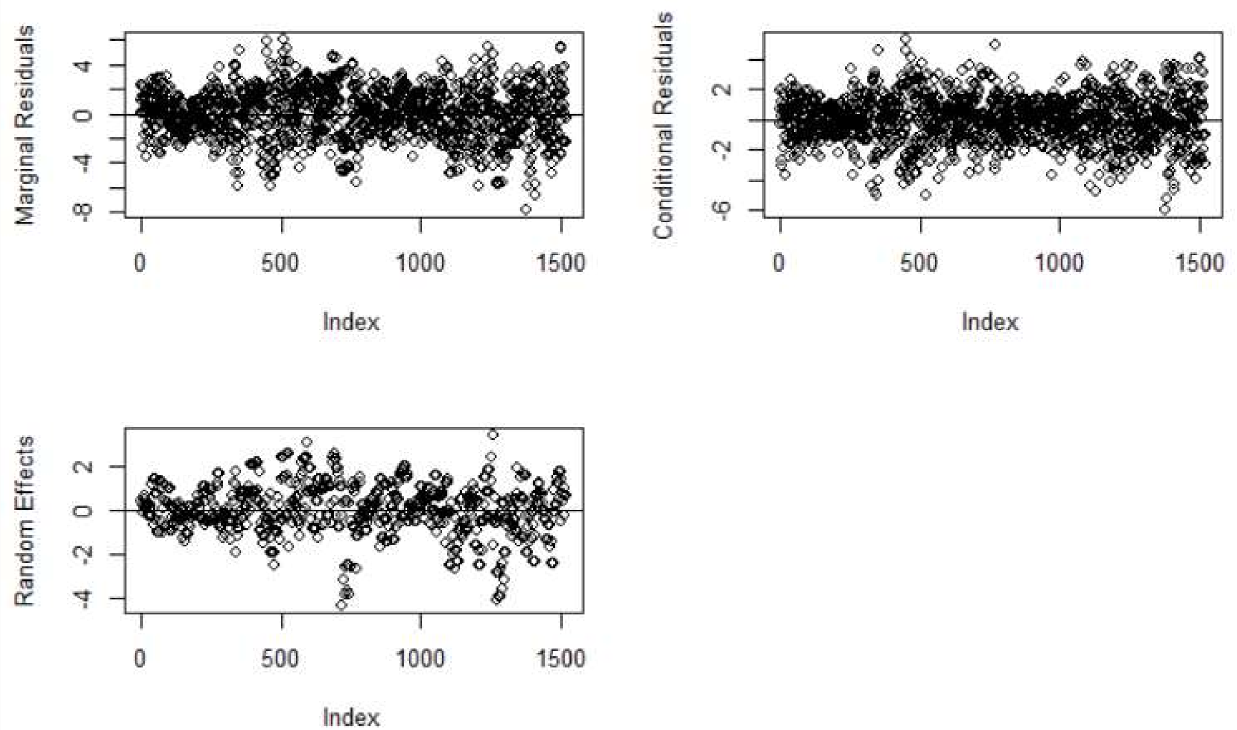


Figure 11

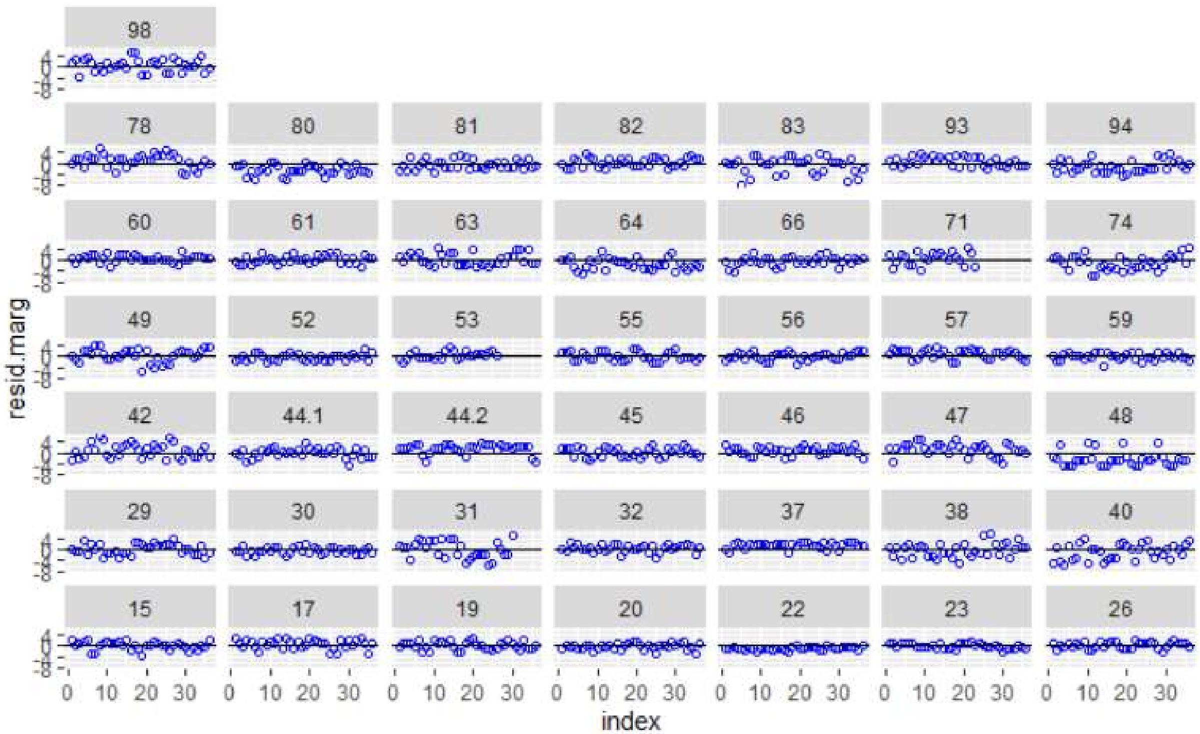


Figure 12

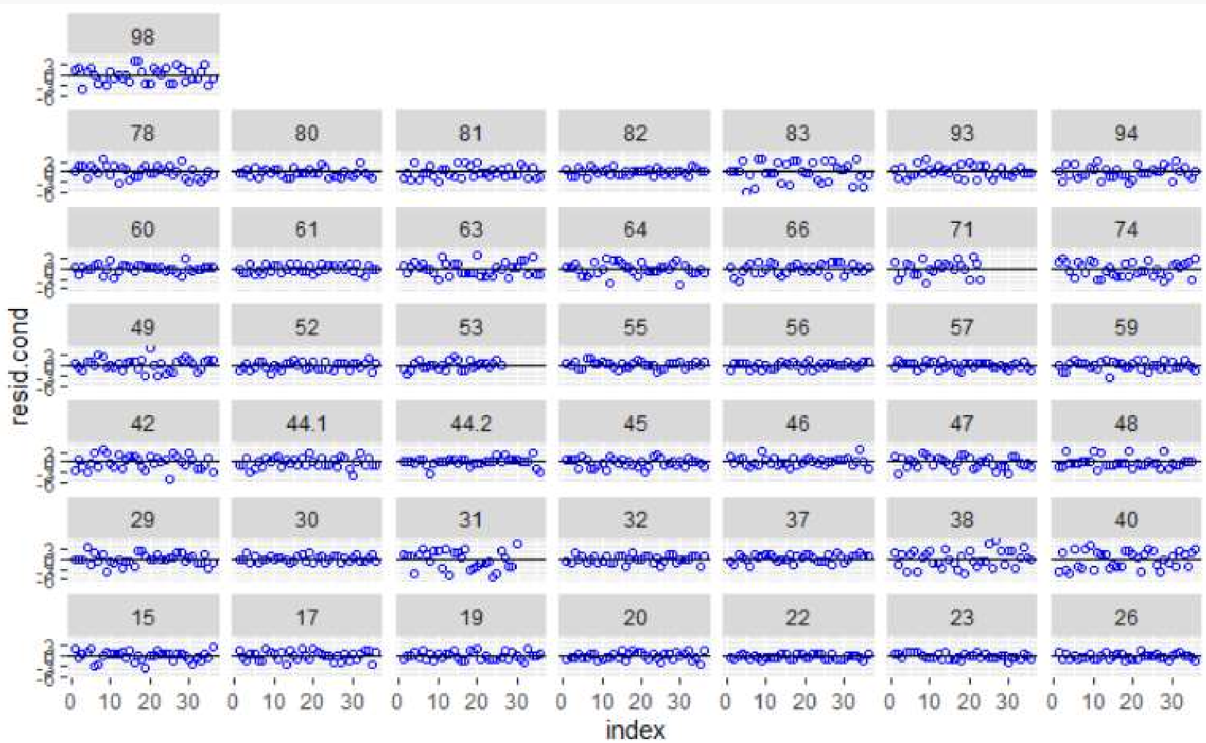


Figure 13

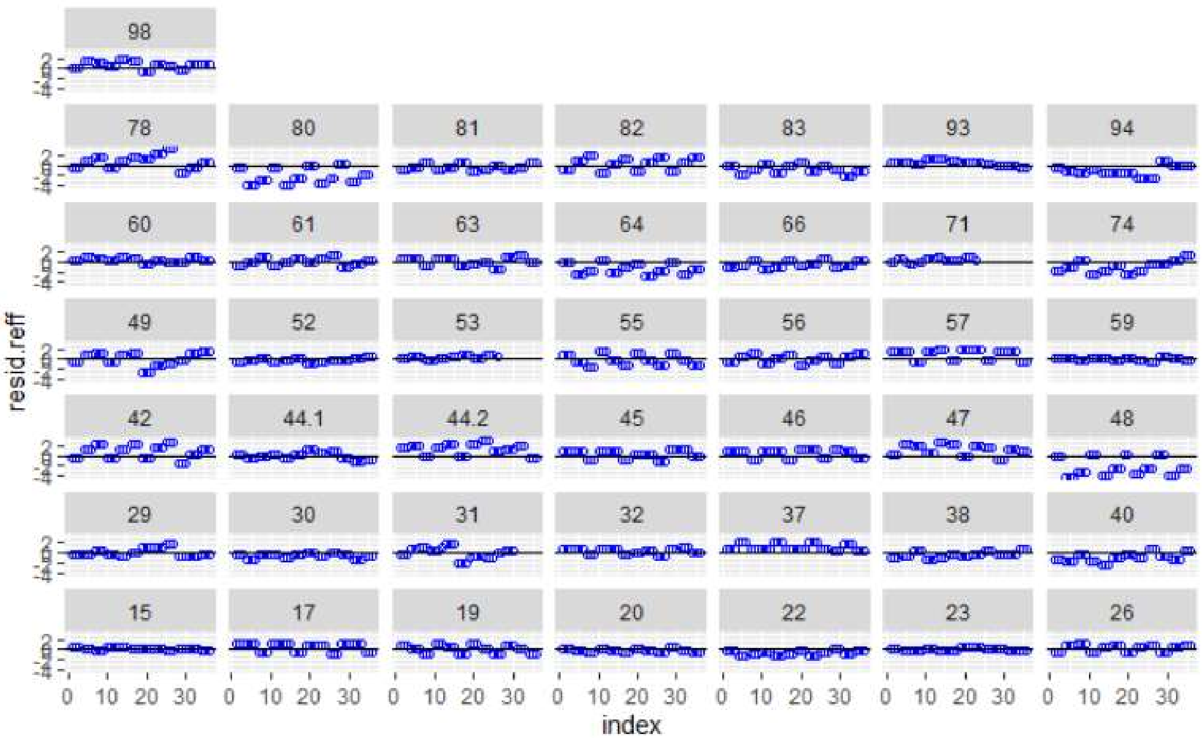

98
78
80
81
82
83
93
94
60
61
63
64
66
71
74
49
52
53
55
56
57
59
42
44.1
44.2
45
46
47
48
29
30
31
32
37
38
40
15
17
19
20
22
23
26
resid.reff
index

Figure 14