Jenny (Jae-Eun) Chang
36763 Hw05

1. *The Three main experimental factors*.
    a. When I regressed Classical against Instrument, Harmony, and Voice each separately, I found most of the variables to be significant at 0.05 significance level.

For Instrument, I got:

```
Call:
lm(formula = Classical ~ Instrument)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.4093 -1.6488 -0.2761  1.5907 11.5907 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       4.27611    0.08058   53.07   <2e-16 ***
Instrumentpiano   1.37267    0.11440   12.00   <2e-16 ***
Instrumentstring  3.13318    0.11371   27.55   <2e-16 ***
```

For Voice, I got:

```
Call:
lm(formula = Classical ~ Voice)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.0436 -2.0436  0.3256  2.3256 12.9564 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.04356    0.09199  65.699  < 2e-16 ***
Voicepar3rd -0.41163    0.13013  -3.163  0.00158 ** 
Voicepar5th -0.36916    0.13005  -2.839  0.00457 ** 
```

For Harmony, I got:

```
Call:
lm(formula = Classical ~ Harmony)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3564 -2.3564  0.3622  2.3622 13.3622 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)     6.3564     0.1058  60.096  < 2e-16 ***
HarmonyI-IV-V  -0.7715     0.1494  -5.164 2.61e-07 ***
HarmonyI-V-IV  -0.8010     0.1496  -5.355 9.36e-08 ***
HarmonyIV-I-V  -0.7186     0.1495  -4.808 1.62e-06 ***
```

To see whether every variable should be included in the model, I compared the model that had only two of the variable versus the model with all of the variables.

First, I tried to see if including **Harmony** is better than not including it and, indeed, ANOVA indicates that the full model is better.

```
Model 1: Classical ~ Instrument
Model 2: Classical ~ Instrument + Voice
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq       F    Pr(>F)
1   2490 13467
2   2488 13381  2    85.603  8.1146 0.0003071 ***
3   2485 13108  3   273.650 17.2934 4.107e-11 ***
```

```
Model 1: Classical ~ Voice
Model 2: Classical ~ Instrument + Voice
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq       F    Pr(>F)
1   2490 17510
2   2488 13381  2    4128.3 391.339 < 2.2e-16 ***
3   2485 13108  3     273.6  17.293 4.107e-11 ***
```

Second, I tried to see if including **Voice** is better than not including it, and, indeed, ANOVA indicates that including it is better.

```
Model 1: Classical ~ Instrument
Model 2: Classical ~ Instrument + Harmony
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq       F    Pr(>F)
1   2490 13467
2   2487 13193  3    273.61 17.2911 4.121e-11 ***
3   2485 13108  2     85.64  8.1181 0.0003061 ***
```

```
Model 1: Classical ~ Harmony
Model 2: Classical ~ Instrument + Harmony
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq        F    Pr(>F)
1   2489 17320
2   2487 13193  2    4127.1 391.2243 < 2.2e-16 ***
3   2485 13108  2      85.6   8.1181 0.0003061 ***
```

Third, I tried to see if including **Instrument** is better than not including it, and, indeed, ANOVA indicates that including it is better.

```
Model 1: Classical ~ Harmony
Model 2: Classical ~ Harmony + Voice
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq        F    Pr(>F)
1   2489 17320
2   2487 17235  2      85.2   8.0779 0.0003185 ***
3   2485 13108  2    4127.6 391.2645 < 2.2e-16 ***
```

```
Model 1: Classical ~ Voice
Model 2: Classical ~ Harmony + Voice
Model 3: Classical ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq       F    Pr(>F)    
1   2490 17510                                   
2   2487 17235  3     274.4  17.343 3.823e-11 ***
3   2485 13108  2    4127.6 391.264 < 2.2e-16 ***
```

When I regressed all three together, I got:

```
Call:
lm(formula = Classical ~ Instrument + Harmony + Voice)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8718 -1.7137 -0.0297  1.7576 11.4766 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)        5.1092     0.1301  39.271  < 2e-16 ***
Instrumentpiano    1.3736     0.1130  12.158  < 2e-16 ***
Instrumentstring   3.1331     0.1123  27.899  < 2e-16 ***
HarmonyI-IV-V     -0.7691     0.1301  -5.913 3.83e-09 ***
HarmonyI-V-IV     -0.8002     0.1302  -6.144 9.33e-10 ***
HarmonyIV-I-V     -0.7190     0.1301  -5.525 3.63e-08 ***
Voicepar3rd       -0.4125     0.1127  -3.660 0.000258 ***
Voicepar5th       -0.3706     0.1126  -3.290 0.001016 ** 
```

Then, using the full model, I analyzed the effects of each variable and each dummy variable for the different variables was significant at the 0.05 significance level. For **Instrument**, the classical rating scores for piano and string are higher than that of guitar by 1.374 and 3.133, respectively. For **Harmony**, the classical rating scores for I-IV-V progression, I-V-IV progression, and IV-I-V progression are all lower than that for I-V-vi by 0.769, 0.800, and 0.719, respectively. For **Voice**, the classical rating scores are lower for parallel 3rd and parallel 5th than that for contrary motion by 0.412 and 0.371, respectively.

b.

i. For each student j, there are 36 ratings. So j ranges from 1 to 70 and i ranges from 1 to 36.

$$Classical_i = \alpha_{0j[i]} + \alpha_1 Instrument_i + \alpha_2 Harmony_i + \alpha_3 Voice_i + \epsilon_i, \quad \epsilon_i \sim^{iid} N(0, \sigma^2)$$

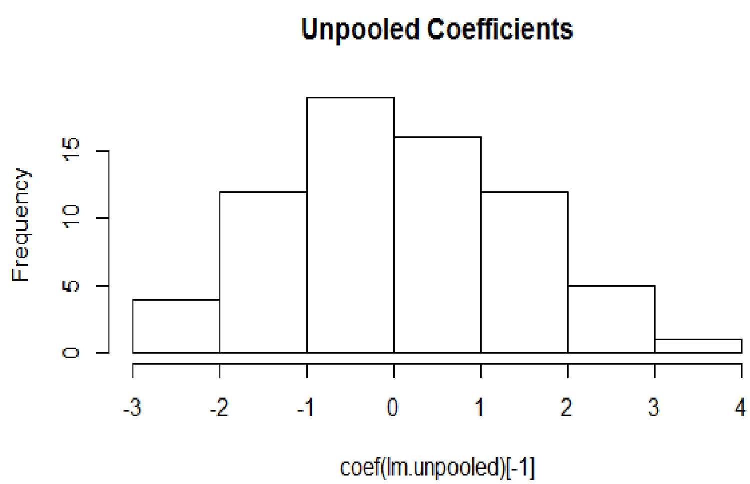
$$\alpha_{0j} = \beta_0 + \eta_j, \quad \eta_j \sim^{iid} N(0, \tau^2)$$

ii. I will provide two methods to test whether the random intercept is needed in the model. First, I checked to see if there is a lot of variation in **Classical** among different **Subjects** by regressing **Classical** against **Subjects** and see if **Subjects** explain Classical well by running an ANOVA. And we find that it does – so a totally pooled model is not the best model to use. Also, I found 43 of the coefficients for different subjects to be significant.

```
> contrasts(Subject) <- contr.sum(70)
> lm.unpooled <- lm(Classical ~ Subject)
> anova(lm.unpooled)
Analysis of Variance Table

Response: Classical
            Df  Sum Sq Mean Sq F value    Pr(>F)    
Subject     69  4462.1  64.668  11.931 < 2.2e-16 ***
Residuals 2423 13132.5   5.420
```

As another test for whether I should pool the data or not, I will check how the coefficients are distributed.



The **Subject** means look rather normal. So, it may be better to model them to have a normal distribution than a uniform distribution.

iii.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ (1 | Subject) + Instrument + Harmony + Voice

REML criterion at convergence: 10471.51

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 1.702    1.305
 Residual             3.581    1.892
Number of obs: 2493, groups: Subject, 70

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       5.11470    0.18925   27.03
Instrumentpiano   1.37705    0.09318   14.78
Instrumentstring  3.13161    0.09257   33.83
HarmonyI-IV-V    -0.77096    0.10718   -7.19
HarmonyI-V-IV    -0.80347    0.10731   -7.49
HarmonyIV-I-V    -0.72106    0.10722   -6.73
Voicepar3rd      -0.41507    0.09287   -4.47
Voicepar5th      -0.37439    0.09281   -4.03
```

The coefficient estimates are not very different from our totally pooled model and all estimates have significantly large t-values. For **Instrument**, the classical rating scores for piano and string are higher than that of guitar by 1.377 and 3.132, respectively. For **Harmony**, the classical rating scores for I-IV-V progression, I-V-IV progression, and IV-I-V progression are all lower than that for I-V-vi by 0.771, 0.803, and 0.721, respectively. For **Voice**, the classical rating scores are lower for parallel 3^{rd} and parallel 5^{th} than that for contrary motion by 0.415 and 0.374, respectively.

We also find the tau squared and sigma squared from equation (1bi) to be 1.702 and 3.581, respectively. Therefore, we find that different people do not vary too greatly (with a variance of 1.702). Classical ratings will vary by somewhat due to chance (variance is not too small).

c.

i. I compare the model that has only one random effect for personal bias and the model that has all three new random effect terms by running ANOVA. And from ANOVA, I find that the model that has all three new random effects is significantly better. The AIC, BIC, and DIC are all smaller by more than 2 for the model with the three new random effect terms than the model with only one random effect term.

```
Models:
M0: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
M1: Classical ~ Instrument + Harmony + Voice + (1 |
Subject:Instrument) +
M1:     (1 | Subject:Harmony) + (1 | Subject:Voice)
   Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M0 10 10469 10527 -5224.4    10449
M1 12 10058 10127 -5016.8    10034 415.33      2  < 2.2e-16 ***
```

For the first model:

```
number of obs: 2493, groups: Subject, 70
AIC = 10491.5, DIC = 10426.2
deviance = 10448.9
```

For the model with three new random effects:

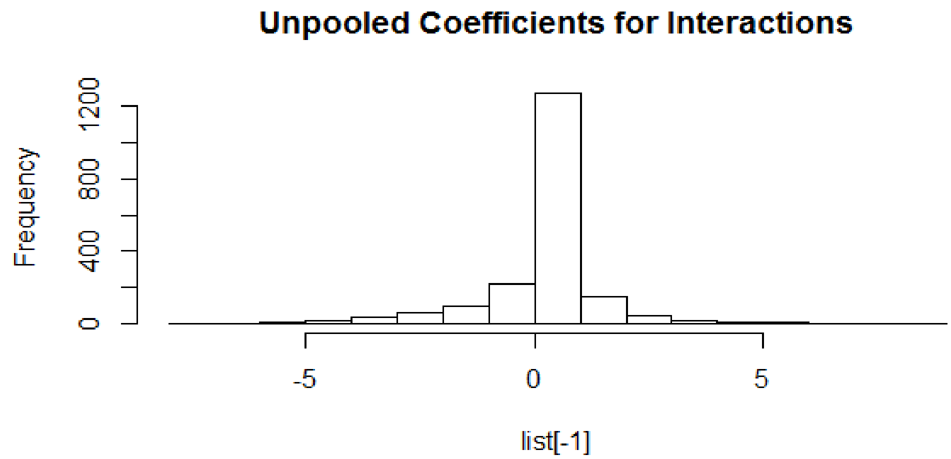
```
number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice,
210; Subject:Instrument, 210
AIC = 10075.5, DIC = 10015.5
deviance = 10033.5
```

I also confirm that the model with three new random effects is better than the pooled model by regressing the following model and getting significant interaction effects between **Subject** and each of **Instrument**, and **Harmony**. Also, the interaction between **Subject** and **Voice** has a p-value that is comparatively small. Therefore, we see that including the three new random effects will be better.

```
> lm.unpooled.subject <- lm(Classical ~ Subject*Instrument + Subject*Harmony +
Subject*Voice)
> anova(lm.unpooled.subject)
Analysis of Variance Table

Response: Classical
                     Df Sum Sq Mean Sq  F value    Pr(>F)    
Subject              69 4462.1   64.67  26.4141 < 2.2e-16 ***
Instrument            2 4118.5 2059.23 841.1025 < 2.2e-16 ***
Harmony               3  275.5   91.82  37.5049 < 2.2e-16 ***
Voice                 2   87.1   43.54  17.7826 2.225e-08 ***
Subject:Instrument  138 2312.2   16.75   6.8436 < 2.2e-16 ***
Subject:Harmony     207 1238.5    5.98   2.4439 < 2.2e-16 ***
Subject:Voice       138  368.3    2.67   1.0902    0.2304    
Residuals          1933 4732.5    2.45                       
```

Also, we are confirmed that estimates for the interactions between **Subject** and each of the **Instrument**, **Harmony**, and **Voice** come from an approximately normal distribution.



ii.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject:Instrument) +      (1
| Subject:Harmony) + (1 | Subject:Voice)

REML criterion at convergence: 10051.51

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.44307  0.6656
 Subject:Voice      (Intercept) 0.02809  0.1676
 Subject:Instrument (Intercept) 2.19850  1.4827
 Residual                       2.43753  1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210;
Subject:Instrument, 210

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       5.11169    0.21443  23.839
Instrumentpiano   1.36384    0.26232   5.199
Instrumentstring  3.12836    0.26203  11.939
HarmonyI-IV-V    -0.77063    0.14316  -5.383
HarmonyI-V-IV    -0.80086    0.14325  -5.590
HarmonyIV-I-V    -0.71445    0.14320  -4.989
Voicepar3rd      -0.40699    0.08174  -4.979
Voicepar5th      -0.37084    0.08168  -4.540
```

The coefficient estimates are not very different from our totally pooled model and all estimates have significantly large t-values. For **Instrument**, the classical rating scores for piano and string are higher than that of guitar by 1.364 and 3.128, respectively. For **Harmony**, the classical rating scores for I-IV-V progression, I-V-IV progression, and IV-I-V progression are all lower than that for I-V-vi by 0.771, 0.801, and 0.714, respectively. For **Voice**, the classical rating scores are lower for parallel 3^{rd} and parallel 5^{th} than that for contrary motion by 0.407 and 0.371, respectively.

Random effect for **Harmony** for each subject has variance 0.443, random effect for **Voice** for each subject has variance 0.028, random effect for **Instrument** for each subject has variance 2.199, and variance for **Classical** is 2.438. Therefore, whether someone is more strongly biased against judging some music to be classical than another person depending on **Harmony** and **Voice** are not very large. However, one's bias towards judging music to be classical depending on the type of instrument varies quite a bit for a person to person. Classical rating will vary by some degree due to chance by a variance of 2.438.

iii.

$$Classical_i = \alpha_{0j[i]} + \alpha'_{0j[i]} + \alpha''_{0j[i]} + \alpha_1 Instrument_i + \alpha_2 Harmony_i + \alpha_3 Voice_i + \epsilon_i, \quad \epsilon \sim^{iid} N(0, \sigma^2)$$

$$\alpha_{0j} = \beta_{00} + \beta_{01} Instrument_j + \eta_j, \quad \eta_j \sim^{iid} N(0, \tau^2)$$

$$\alpha'_{0j} = \beta_{03} + \beta_{04} Harmony_j + \eta'_j, \quad \eta'_j \sim^{iid} N(0, \tau^{2\prime})$$

$$\alpha''_{0j} = \beta_{05} + \beta_{06} Harmony_j + \eta''_j, \quad \eta''_j \sim^{iid} N(0, \tau^{2\prime\prime})$$

2. *Individual Covariates*
   a. For adding covariates, I considered adding the following variables in my model: **OMSI**, **OMSI*Harmony, OMSI*Voice**, **Selfdeclare, X16.minus.17**, **Instrument *Instr.minus.Notes**, **Voice*Inst.minus.Notes**, **Harmony*Inst.minus.Notes**, **PachListen, ClsListen, KnowRob,** and **KnowAxis.** The reasons I chose to include the variables above are:
      i. **OMSI**: How much musical knowledge one has may significantly affect one's ability to distinguish voice leading and is more likely to know whether certain voice leading is more likely to occur in classical music versus popular music. Therefore, I include this variable along with two interaction terms: **OMSI*Voice** and **OMSI*Harmony**. This also is very much likely to be highly correlated with other variables included in our data such as **CollegeMusic, NoClass, APTheory, Composing, PianoPlay, GuitarPlay, X1stInstr,** and **X2ndInstr.**
      ii. **Selfdeclare:** This is a variable of interest to the person who ran the experiment.
      iii. **X16.minus.17**: If one scores higher on this measure then one is much more likely to be accurate on its rating. Therefore, this effect will most likely be a significant factor in our model.
      iv. **Inst.minus.Notes**: If one concentrates more on voice leading or instrument one may be more heavily influenced by **Instrument** or **Voice**. So interaction terms, **Instrument*Instr.minus.Notes**, **Harmony*Instr.minus.Notes**, and **Voice*Instr.minus.Notes**, should be included.
      v. **PachListen**: Knowing Pachelbel's Cannon may influence someone to rate music with I-V-vi progression as classical, as predicted in one of the hypotheses by the professor.
      vi. **ClsListen**: People who listen to a lot of Classical music will be more able to identify some of the classical music that is played in the experiment.
      vii. **X1990s200s**: People who listen to a lot of popular music will be more able to identify some of the popular music that is played in the experiment.
      viii. **KnowRob** and **KnowAxis**: Knowing these two things may influence someone's ability to identify certain voice leading in the musical pieces and convince someone to rate differently.

ix. And if we regress **Classical** against all of the variables we get the outcome below and find that certain variables are not significant at 0.05 significance level: **KnowRob, KnowAxis, OMSI*Voice,** and **Voice*Instr.minus.Notes**.

```
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          3.153e+00  3.556e-01   8.868  < 2e-16 ***
OMSI                                 2.642e-03  5.702e-04   4.633 3.83e-06 ***
Voicepar3rd                         -4.591e-01  1.835e-01  -2.502 0.012441 *  
Voicepar5th                         -3.027e-01  1.835e-01  -1.649 0.099248 .  
HarmonyI-IV-V                       -7.364e-01  2.120e-01  -3.474 0.000524 ***
HarmonyI-V-IV                       -7.959e-01  2.120e-01  -3.755 0.000178 ***
HarmonyIV-I-V                       -8.265e-01  2.120e-01  -3.899 9.96e-05 ***
X16.minus.17                        -1.515e-01  1.700e-02  -8.915  < 2e-16 ***
Instrumentpiano                      1.344e+00  1.239e-01  10.853  < 2e-16 ***
Instrumentstring                     3.037e+00  1.233e-01  24.641  < 2e-16 ***
Instr.minus.Notes                   -1.994e-01  7.772e-02  -2.565 0.010384 *  
PachListen                           3.279e-01  5.829e-02   5.625 2.10e-08 ***
ClsListen                            3.083e-01  3.229e-02   9.550  < 2e-16 ***
KnowRob                             -8.232e-03  3.249e-02  -0.253 0.800000    
KnowAxis                             5.250e-02  2.687e-02   1.954 0.050853 .  
Selfdeclare                         -2.839e-01  6.930e-02  -4.096 4.36e-05 ***
OMSI:Voicepar3rd                     1.541e-04  5.042e-04   0.306 0.759895    
OMSI:Voicepar5th                    -2.782e-05  5.042e-04  -0.055 0.955997    
OMSI:HarmonyI-IV-V                  -1.349e-03  5.823e-04  -2.317 0.020598 *  
OMSI:HarmonyI-V-IV                  -1.113e-03  5.831e-04  -1.908 0.056535 .  
OMSI:HarmonyIV-I-V                  -6.317e-04  5.823e-04  -1.085 0.278079    
Instrumentpiano:Instr.minus.Notes    1.022e-01  6.475e-02   1.579 0.114532    
Instrumentstring:Instr.minus.Notes   3.753e-01  6.441e-02   5.827 6.54e-09 ***
Voicepar3rd:Instr.minus.Notes        5.672e-02  6.718e-02   0.844 0.398640    
Voicepar5th:Instr.minus.Notes       -1.499e-03  6.718e-02  -0.022 0.982202    
HarmonyI-IV-V:Instr.minus.Notes      2.413e-01  7.758e-02   3.111 0.001891 ** 
HarmonyI-V-IV:Instr.minus.Notes      2.817e-01  7.762e-02   3.630 0.000291 ***
HarmonyIV-I-V:Instr.minus.Notes      2.135e-01  7.758e-02   2.753 0.005965 ** 
```

Therefore, I tried running the ANOVA to check whether the model that includes these four variables is better than not including them and including these variables does not make the model significantly better. I also checked the AIC and BIC and they were much lower for the model without those variables. Therefore, I choose to use the model with fewer variables. So I include the following variables as fixed effects in my final model: **Harmony, Instrument, Voice, OMSI*Harmony, X16.minus.17, Instrument*Instr.minus.Notes, Harmony*Instr.minus.Notes, PachListen, ClsListen.**

```
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   2095 9719.5                           
2   2089 9695.8  6    23.709 0.8514 0.5302
> AIC(model.new.drop);AIC(model.new)
[1] 9280.372
[1] 9287.201
> BIC(model.new.drop);BIC(model.new)
[1] 9410.5
[1] 9451.276
```

b. I compared six different models each model had one more of the random effect term of the following random effect terms: (1|**Subject:OMSI**), (1|**Subject:X16.minus.17**), (1|**Subject:Instr.minus.Notes**), (1|**Subject:PachListen**), (1|**Subject:Selfdeclare**) and (1|**Subject:ClsListen**). And when I ran an ANOVA to determine including which random effect term is best I found that the model should include one new random effect term, (1|**Subject:OMSI**) . Also, the AI C and BIC is the lowest for the model that includes only those two of the random effects.

```
       Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)    
Mnew.0 26 8483.5 8630.6 -4215.8   8431.5                              
Mnew.1 27 8467.9 8620.7 -4207.0   8413.9 17.5672      1  2.773e-05 ***
Mnew.2 27 8469.6 8622.3 -4207.8   8415.6  0.0000      0     1.0000    
Mnew.3 29 8471.9 8636.0 -4207.0   8413.9  1.6175      2     0.4454    
Mnew.4 30 8473.9 8643.7 -4207.0   8413.9  0.0000      1     1.0000    
Mnew.5 31 8475.9 8651.3 -4207.0   8413.9  0.0000      1     1.0000    
Mnew.6 32 8477.9 8659.0 -4207.0   8413.9  0.0000      1     1.0000    
```

c. So, from the model I decided upon in the previous parts, I get the following result:

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.4002   0.6326  
 Subject:Voice      (Intercept) 0.0194   0.1393  
 Subject:Instrument (Intercept) 1.1360   1.0658  
 Subject:OMSI       (Intercept) 0.8622   0.9286  
 Residual                       2.3989   1.5488  
Number of obs: 2117, groups: Subject:Harmony, 236; Subject:Voice, 177;
Subject:Instrument, 177; Subject:OMSI, 59

Fixed effects:
                                   Estimate Std. Error t value
(Intercept)                       3.1102166  1.0257034   3.032
HarmonyI-IV-V                    -0.7278476  0.2408851  -3.022
HarmonyI-V-IV                    -0.7852393  0.2408899  -3.260
HarmonyIV-I-V                    -0.8190131  0.2408850  -3.400
Instrumentpiano                   1.3347634  0.2292858   5.821
Instrumentstring                  3.0399605  0.2289874  13.276
Selfdeclare                      -0.2753483  0.2290704  -1.202
Voicepar3rd                      -0.3766901  0.0863721  -4.361
Voicepar5th                      -0.3096980  0.0863721  -3.586
OMSI                              0.0026911  0.0011611   2.318
X16.minus.17                     -0.1549236  0.0558215  -2.775
Instr.minus.Notes                -0.1664613  0.1277916  -1.303
PachListen                        0.3341564  0.1917738   1.742
ClsListen                         0.3085255  0.1042041   2.961
HarmonyI-IV-V:OMSI               -0.0013780  0.0006603  -2.087
HarmonyI-V-IV:OMSI               -0.0011755  0.0006607  -1.779
HarmonyIV-I-V:OMSI               -0.0006449  0.0006603  -0.977
Instrumentpiano:Instr.minus.Notes 0.1083222  0.1198679   0.904
Instrumentstring:Instr.minus.Notes 0.3747648 0.1197054   3.131
HarmonyI-IV-V:Instr.minus.Notes   0.2411751  0.0881400   2.736
HarmonyI-V-IV:Instr.minus.Notes   0.2840373  0.0881571   3.222
HarmonyIV-I-V:Instr.minus.Notes   0.2122328  0.0881397   2.408
```

I also tried dropping the variables that were not significant (I consider those with t-value less than 1.96 to be insignificant) such as **PachListen** and **Selfdeclare**. However, dropping the variables does not cause much difference (AIC are virtually the same) between not dropping them so I decide to drop them for my final model on classical ratings.

```
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Mnew.1.new 25 8469.5 8610.9 -4209.7   8419.5
Mnew.1     27 8468.2 8621.0 -4207.1   8414.2 5.2497      2    0.07245 .
```

Therefore, we fit the following model:

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.39878  0.6315
 Subject:Voice      (Intercept) 0.01941  0.1393
 Subject:Instrument (Intercept) 1.13462  1.0652
 Subject:OMSI       (Intercept) 0.95728  0.9784
 Residual                       2.39890  1.5488
Number of obs: 2117, groups: Subject:Harmony, 236; Subject:Voice,
177; Subject:OMSI, 59

Fixed effects:
                                   Estimate Std. Error t value
(Intercept)                       4.4829774  0.7560895   5.929
HarmonyI-IV-V                    -1.4688873  0.4509396  -3.257
HarmonyI-V-IV                    -1.7652763  0.4509649  -3.914
HarmonyIV-I-V                    -1.5461947  0.4509391  -3.429
Instrumentpiano                   1.0094900  0.5264396   1.918
Instrumentstring                  1.8011571  0.5253875   3.428
Voicepar3rd                      -0.3766847  0.0863724  -4.361
Voicepar5th                      -0.3097050  0.0863724  -3.586
OMSI                              0.0020119  0.0008225   2.446
X16.minus.17                     -0.1634436  0.0570749  -2.864
Instr.minus.Notes                -0.0685080  0.0522386  -1.311
ClsListen                         0.3369697  0.1351018   2.494
HarmonyI-IV-V:OMSI               -0.0014165  0.0006580  -2.153
HarmonyI-V-IV:OMSI               -0.0011733  0.0006584  -1.782
HarmonyIV-I-V:OMSI               -0.0006452  0.0006580  -0.981
Instrumentpiano:Instr.minus.Notes  0.0403669  0.0483553   0.835
Instrumentstring:Instr.minus.Notes 0.1511105  0.0482848   3.130
HarmonyI-IV-V:Instr.minus.Notes   0.0925446  0.0354507   2.611
HarmonyI-V-IV:Instr.minus.Notes   0.1186013  0.0354586   3.345
HarmonyIV-I-V:Instr.minus.Notes   0.0881575  0.0354505   2.487
```

I judge the coefficients with t-value greater than 1.96 or less than -1.96 to be significant predictors. Therefore, I find that the classical rating for **Harmony** progression I-V-vi is higher than Harmony I-IV-V, I-V-IV, and IV-I-V, by 1.469, 1.769, and 1.546, respectively. Also, for **Instrument**, the classical rating scores for strings are higher than that of guitar by 1.801. For **Voice**, the classical rating scores are lower for parallel 3rd and parallel 5th than that for contrary motion by 0.377 and 0.31, respectively. An increase in **OMSI** the classical ratings go up by 0.002. For an increase in **Instr.minus.Notes** classical ratings go down by 0.069. For **ClsListen**, classical ratings increases by 0.337 per an increase in **ClsListen**. For an increase in **X16.minus.17** by one point, the classical rating score decreases by 0.155. For an increase in **ClsListen** by one point, the classical rating score increases by 0.309. For an interaction between **HarmonyI-IV-V:OMSI**, for a piece that had a harmonic progression I-IV-V, an additional increase in **OMSI** by one point is going to result in lower classical rating score by 0.001. For a string instrument, an additional increase in **Instr.minus.Notes**, will increase classical rating score by 0.151. Also, for the harmonic progression I-IV-V, an additional increase in **Instr.minus.Notes** is going to increase classical ratings by 0.093. For the harmonic progression I-V-IV, an additional increase in **Instr.minus.Notes** is going to increase classical ratings by 0.119. And finally, for the harmonic progression I-V-IV, an additional increase in **Instr.minus.Notes** is going to increase classical ratings by 0.088.

The random effects for **Harmony, Instrument, Voice,** and **OMSI** have variance 0.399, 1.136, 0.019, and 0.862, respectively. The variances are quite small, indicating that there is not much personal bias due to these different variables. It seems that **Instrument** has the strongest effect on a person's tendency to rate some music as classical. Also, the residuals of the fixed effects regression had a variance of 2.399. The scores do not seem to vary too much due to random variation.

3. When I ran the model in the second question **Selfdeclare** variable was not significant when we included the random effects. Therefore, I tried to see if there may be a significant interaction with **Selfdeclare** and another variable. I found only the following interaction to be significant: **Selfdeclare*Harmony**.

```
Fixed effects:
                                     Estimate Std. Error t value
(Intercept)                         3.6319372  1.2374004   2.935
HarmonyI-IV-V                      -0.1728025  0.5545689  -0.312
HarmonyI-V-IV                      -0.5793357  0.5546632  -1.044
HarmonyIV-I-V                      -0.1797968  0.5545663  -0.324
Selfdeclare                         0.3028251  0.4387669   0.690
Instrumentpiano                     1.5072495  0.8134633   1.853
Instrumentstring                    2.4505955  0.8133160   3.013
Voicepar3rd                        -0.3690479  0.2097158  -1.760
Voicepar5th                        -0.2978135  0.2097158  -1.420
OMSI                                0.0013069  0.0012385   1.055
X16.minus.17                        0.0896495  0.1509339   0.594
Instr.minus.Notes                  -0.0511354  0.0523190  -0.977
ClsListen                           0.1235874  0.3448722   0.358
HarmonyI-IV-V:Selfdeclare          -0.7102159  0.1926210  -3.687
HarmonyI-V-IV:Selfdeclare          -0.6496076  0.1926444  -3.372
HarmonyIV-I-V:Selfdeclare          -0.7489225  0.1926203  -3.888
Selfdeclare:Instrumentpiano        -0.1528622  0.1896836  -0.806
Selfdeclare:Instrumentstring       -0.1988735  0.1896054  -1.049
Selfdeclare:Voicepar3rd            -0.0030732  0.0746304  -0.041
Selfdeclare:Voicepar5th            -0.0045081  0.0746305  -0.060
HarmonyI-IV-V:OMSI                  0.0012194  0.0009497   1.284
HarmonyI-V-IV:OMSI                  0.0012400  0.0009503   1.305
HarmonyIV-I-V:OMSI                  0.0021326  0.0009497   2.246
Selfdeclare:X16.minus.17           -0.1034734  0.0600997  -1.722
Instrumentpiano:Instr.minus.Notes   0.0296233  0.0503399   0.588
Instrumentstring:Instr.minus.Notes  0.1369885  0.0502908   2.724
HarmonyI-IV-V:Instr.minus.Notes     0.0803375  0.0338911   2.370
HarmonyI-V-IV:Instr.minus.Notes     0.1073612  0.0339004   3.167
HarmonyIV-I-V:Instr.minus.Notes     0.0753372  0.0338908   2.223
Selfdeclare:ClsListen               0.1026436  0.1256520   0.817
```

4. *Classical vs. Popular*
   a. Unlike for **Classical,** the influence of some of the variables is not significant on **Popular** ratings. As we can see below, **Harmony** does not seem to be significant for factor level. Also, we find that **Voice** is not a significant predictor in our model.

```
Call:
lm(formula = Popular ~ Instrument + Harmony + Voice)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.7218 -1.7026  0.2008  1.4691 13.2248 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       6.31434    0.12784  49.391   <2e-16 ***
Instrumentpiano  -0.95200    0.11102  -8.575   <2e-16 ***
Instrumentstring -2.61173    0.11035 -23.667   <2e-16 ***
HarmonyI-IV-V     0.26829    0.12782   2.099   0.0359 *  
HarmonyI-V-IV     0.24425    0.12797   1.909   0.0564 .  
HarmonyIV-I-V     0.08265    0.12787   0.646   0.5181    
Voicepar3rd       0.16859    0.11075   1.522   0.1281    
Voicepar5th       0.16326    0.11068   1.475   0.1403    
```

Therefore, we should definitely include **Instrument** in our model but we need to check if adding **Harmony** or **Voice,** or both will give a better model. So, we ran ANOVA to compare to see if adding **Harmony** would produce a better model and find that that is not true.

harmony and voice need to be in the model because they are design factors

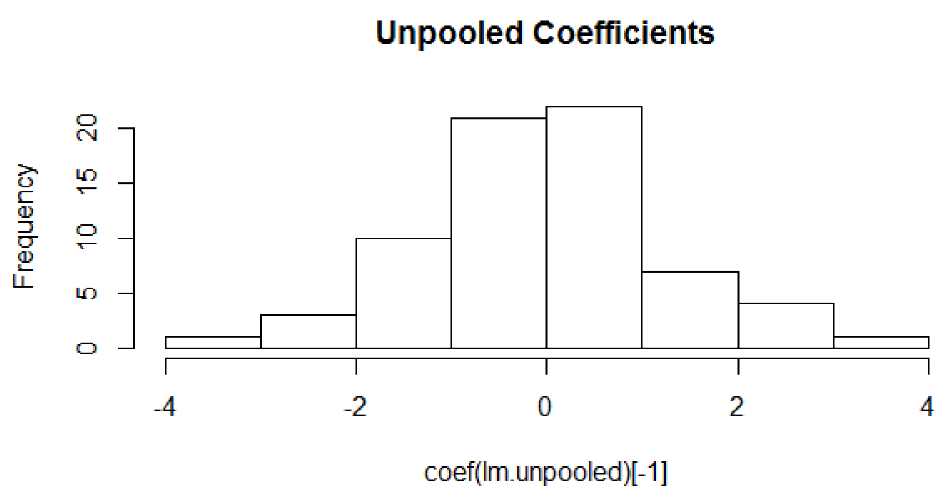
```
Model 1: Popular ~ Instrument
Model 2: Popular ~ Instrument + Harmony
Model 3: Popular ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   2490 12703
2   2487 12672  3    31.119 2.0367 0.1067
3   2485 12656  2    15.263 1.4984 0.2237
```

We also ran ANOVA to compare to see if adding **Voice** would be better and find that that is not true either.

```
Model 1: Popular ~ Instrument
Model 2: Popular ~ Instrument + Voice
Model 3: Popular ~ Instrument + Harmony + Voice
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   2490 12703
2   2488 12688  2    15.291 1.5011 0.2231
3   2485 12656  3    31.092 2.0349 0.1069
```

Therefore, we have sufficient evidence not to add **Harmony** or **Voice.** Now, we see if any random effect should be included in the model.

When we regress **Popular** against **Subject** we find 34 coefficients to be significant, meaning **Subject** does have quite a large influence on popular music ratings. We also find that the coefficients are normally distributed. Therefore, we may consider partially pooled coefficients.



nice to check

However, there may be personal bias on a person's likeliness to rate something as popular music depending on which instrument was used. Therefore, I also consider adding a random effect term (1|**Subject:Instrument**) and compare the two different models with the new random effect term and without. And we find that the model with the new random effect term.

```
Models:
lmer.pop: Popular ~ (1 | Subject) + Instrument
lmer.pop.new: Popular ~ (1 | Subject:Instrument) + Instrument
             Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer.pop      5 10434 10463 -5211.9    10424
lmer.pop.new  5 10167 10196 -5078.7    10157 266.4      0  < 2.2e-16 ***
```

b. Now, we consider adding covariates. I considered adding the variables that I mentioned in problem 2 for the same reasons but without the interaction terms that included **Harmony** and **Voice**. I find that **OMSI, PchListen,** and **KnowAxis** are not significant predictors for popular ratings.

```
Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        5.7220326  0.3513889  16.284  < 2e-16 ***
Instrumentpiano                   -0.2598986  0.2909826  -0.893 0.371865
Instrumentstring                  -1.4861593  0.2887865  -5.146 2.90e-07 ***
OMSI                              -0.0002725  0.0003400  -0.802 0.422871
X16.minus.17                       0.0633951  0.0173719   3.649 0.000269 ***
Instr.minus.Notes                  0.0470589  0.0195340   2.409 0.016078 *
PachListen                        -0.1541724  0.0595000  -2.591 0.009632 **
ClsListen                         -0.0719575  0.0440888  -1.632 0.102807
KnowRob                            0.4787218  0.0805134   5.946 3.21e-09 ***
KnowAxis                           0.1065381  0.0693079   1.537 0.124402
Selfdeclare                        0.1750080  0.0701962   2.493 0.012739 *
Instrumentpiano:Instr.minus.Notes -0.0776627  0.0266841  -2.910 0.003647 **
Instrumentstring:Instr.minus.Notes -0.1383792 0.0265368  -5.215 2.02e-07 ***
```

And ANOVA confirms that dropping these variables do not cause much difference in the model.

```
Model 1: Popular ~ Instrument + X16.minus.17 + Instrument * Instr.minus.Notes +
    PachListen + KnowRob + Selfdeclare
Model 2: Popular ~ Instrument + OMSI + X16.minus.17 + Instrument * Instr.minus.Notes +
    PachListen + ClsListen + KnowRob + KnowAxis + Selfdeclare
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   2107 10201
2   2104 10176  3    24.778 1.7077 0.1634
```

By comparing AIC and BIC we see that the best model is the one that includes the random effects: (1|**Subject:Instrument**) and (1|**Subject:X16.minus17**).

what's teh substantive motivation for including this?

```
      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
pop.1 12 8666.9 8734.8 -4321.5   8642.9
pop.2 13 8629.9 8703.4 -4301.9   8603.9 39.038      1  4.155e-10 ***
pop.3 14 8631.9 8711.1 -4301.9   8603.9  0.000      1          1
pop.4 15 8633.9 8718.7 -4301.9   8603.9  0.000      1          1
pop.5 16 8635.9 8726.4 -4301.9   8603.9  0.000      1          1
pop.6 17 8637.9 8734.1 -4301.9   8603.9  0.000      1          1
```

However, when I ran the model there were many variables which were not significant (I consider variables with t-value greater than 1.96 to be significant).

```
Random effects:
 Groups               Name        Variance Std.Dev.
 Subject:Instrument   (Intercept) 0.9027   0.9501
 Subject:X16.minus.17 (Intercept) 1.2098   1.0999
 Residual                         2.8906   1.7002
Number of obs: 2117, groups: Subject:Instrument, 177; Subject:X16.minus.17, 59

Fixed effects:
                                   Estimate Std. Error t value
(Intercept)                         5.68160    1.02615   5.537
Instrumentpiano                    -0.29280    0.48765  -0.600
Instrumentstring                   -1.48800    0.48636  -3.059
X16.minus.17                        0.05902    0.05947   0.992
Instr.minus.Notes                   0.05407    0.04736   1.142
PachListen                         -0.15747    0.19973  -0.788
KnowRob                             0.49400    0.24369   2.027
Selfdeclare                         0.11871    0.15879   0.748
Instrumentpiano:Instr.minus.Notes  -0.07513    0.04478  -1.678
Instrumentstring:Instr.minus.Notes -0.13829    0.04470  -3.094
```

I wanted to see again if dropping the insignificant variables is okay to do for the better model and found that dropping **Selfdeclare, PachListen,** and **X16.minus.Notes** do not change the fit of the model very much.

```
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
pop.2.drop 10 8626.7 8683.3 -4303.4   8606.7
pop.2      13 8629.9 8703.4 -4301.9   8603.9 2.8469      3     0.4158
```

Therefore, I decide to go with the more concise model and consider my final model to be the following.

```
Random effects:
 Groups               Name        Variance Std.Dev.
 Subject:Instrument   (Intercept) 0.9027   0.9501
 Subject:X16.minus.17 (Intercept) 1.1992   1.0951
 Residual                         2.8906   1.7002
Number of obs: 2117, groups: Subject:Instrument, 177; Subject:X16.minus.17, 59

Fixed effects:
                                   Estimate Std. Error t value
(Intercept)                         5.48323    0.64341   8.522
Instrumentpiano                    -0.29350    0.48765  -0.602
Instrumentstring                   -1.48803    0.48636  -3.060
Instr.minus.Notes                   0.05278    0.04624   1.141
KnowRob                             0.52115    0.22948   2.271
Instrumentpiano:Instr.minus.Notes  -0.07508    0.04478  -1.676
Instrumentstring:Instr.minus.Notes -0.13829    0.04470  -3.094
```

As I did before, we will consider the coefficients with t-value greater than 1.96 or less than -1.96 to be significant. We find that string instruments are likely to score lower on popular ratings than guitar music by 1.488. Also, **KnowRob** is a significant predictor and knowing Rob's Rank will increase the popular rating score by 0.521. Also, the interaction between string instrument and **Instr.minus.Notes** is significant. For music that includes strings, for an increase in **Instr.minus.Notes** popular ratings decrease by 0.138.

We also see that the random effects for **Subject:Instrument** and **Subject:X16.minus.17** are 0.903 and 1.199, respectively. Also, the overall variance is equal to 2.891. Therefore, we do not seem that the random effects do not vary very much from person to person. However, there is bigger variation to be considered among scores themselves due to random variation.

c. As I found before, **Selfdeclare** is not a significant variable in my model. Even when I included all the possible interaction effects with **Selfdeclare** I found no interaction to be significant.

you didn't dichotomize self-declare, as instructed.

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	7.158640	2.563596	2.792
Instrumentpiano	-0.427615	0.756282	-0.565
Instrumentstring	-1.777564	0.756102	-2.351
Selfdeclare	-0.487540	0.968255	-0.503
X16.minus.17	0.141401	0.160452	0.881
Instr.minus.Notes	0.057871	0.049418	1.171
PachListen	-0.587058	0.649882	-0.903
KnowRob	0.484397	0.836031	0.579
Instrumentpiano:Selfdeclare	0.041131	0.176361	0.233
Instrumentstring:Selfdeclare	0.088651	0.176265	0.503
Selfdeclare:X16.minus.17	-0.036170	0.063579	-0.569
Instrumentpiano:Instr.minus.Notes	-0.072174	0.046813	-1.542
Instrumentstring:Instr.minus.Notes	-0.131988	0.046752	-2.823
Selfdeclare:PachListen	0.178147	0.254358	0.700
Selfdeclare:KnowRob	-0.004467	0.249367	-0.018

5. *Brief Writeup*.

CLASSICAL MUSIC RATINGS

Harmony is a significant predictor for predicting how classical the musical piece sounds to someone. I found that the chord progression I-V-vi will most likely give the highest rating on how classical a piece sounds. Other progressions I-IV-V, I-V-IV, and IV-I-V will produce a rating lower than the progression I-V-vi by at least 1.5 points.

Instrument is a significant predictor for predicting how classical the musical piece sounds to someone. I found that a piece that contains guitar is least likely to score high on classical ratings. Pieces that contain piano or strings will score higher by at least one point.

Voice is also a significant predictor for predicting how classical the musical piece sounds to someone. I found that a piece that has contrary motion rates highest on classical ratings. Pieces that contain parallel 3rds and parallel 5ths are likely to score lower by at least 0.3 points.

From our data, I also found some other variables to be significant in determining how classical music sounds to a person: **OMSI** (which represents one's musical knowledge and it increases one's score on classical ratings by 0.002)**, X16.minus.17** (which represents a measure of someone's ability to distinguish classical music from popular music and it lowers one's classical ratings by 0.163)**, Instr.minus.Notes** (which indicates whether someone concentrated more on listening to which instrument was used versus focusing more on notes and it decreases classical ratings score by 0.069)**, ClsListen** (which indicates how much someone listens to classical music and it increases classical ratings by 0.337)**, Harmony*OMSI** (for harmonic progression I-IV-V an increase in **OMSI** decreases classical ratings score by 0.001)**, Instrument*Instr.minus.Notes** (for string instruments, an increase in **Instr.minus.Notes** classical ratings increases by 0.151)**,** and **Harmony*Instr.minus.Notes** (for all different levels of harmonic progressions, an increase in **Instr.minus.Notes** increases classical ratings by at least 0.08).

I also noticed that there may be some bias due to the fact that multiple scores were gathered from each person. Personal biases vary with the type of instrument, harmony, voice leading, and musical knowledge. I confirmed from the analysis of fitting different models that different people have different personal biases on these aspects. From analyzing the different random effects, we find that different people don't vary too much on their biases for rating classical music depending on which instrument was used, which voice motion, which chord progression was used, and musical knowledge. But people tended to vary the most for their bias regarding the instrument that was used in the music (detailed analysis is explained in 2a).

POPULAR MUSIC RATINGS

Instrument is a significant predictor for predicting how popular the musical piece sounds to someone. As was the case for classical music ratings, pieces with string instruments and piano were less likely to score higher on popular music ratings than those that included guitar by at least 0.3 points.

Harmony and **Voice** are not significant predictors in predicting whether a piece of music was likely to score higher on popular music ratings.

From our data, I also found other variables to be significant in determining how a piece of music is going to score on popular music ratings: **Instr.music.Notes** (which is not significant by itself but produces a significant interaction effect with **Instrument), KnowRob** (which lowers the popular ratings score by 0.521)**,** and **Instrument*Instr.minus.Notes** (lowers the popular ratings score by at least 0.075).

I found that personal biases vary with the type of instrument and the measure of someone's ability to distinguish classical versus popular music. I found that personal biases do not vary much depending on person for both of these aspects but they do contribute some amount (detailed analysis on variances of random effects on part 3b).

For neither of classical ratings nor popular ratings, I was able to find whether one declares oneself as a musician or non-musician has a significant effect on how someone rates a piece of music as classical or popular.

4: 14
5: 20

34/40