36-763 - Assignment 5

Michael A. Pane

Problem 1

a We start by comparing the full model with all three variables (instrument, harmony, and voice) added into our model. We then compare this model to every combination of the three variables by creating a total of 7 different models. All seven models are not shown but can be seen in the attached R appendix. The full summary output of the model we deemed most appropriate (the full model) is displayed below. Table 1 displays the AIC and BIC for each of the seven models. It is also important to point out that there are 27 NA values in classical ratings and therefore determine to leave them out of the analysis going forward. We may want to check this decision with investigators in future work.

When comparing AIC of the 7 models, we observe that the full model with harmony, instrument, and voice performs the strongest while the BIC for the full model is slightly higher (.6) than the model including harmony and instrument only. We also observe the R^2_{adj} (0.2529) is higher for the full model. We move forward with our analysis with the full model as BIC (which tends to pick simpler models) only slightly prefers the simpler model, and our full model has all 3 variables as being significant. When plotting the diagnostics of the final model (Figure 1), we don't observe a random scatter is the residuals versus fitted values which may be concerning. We also observe one potentially influential outlier. We move forward keeping the point in our analysis but in future work we may want to explore why that point is influential (i.e. coding error) and debate if it should be kept in our analysis.

Table 1: AIC and BIC of the 7 models

	Harm/Instr/Voice	Harm/Instr	Harm/Voice	Instr/Voice	Harm	Voice	Instr
AIC	11230.45	11242.69	11908.94	11275.96	11917.23	11942.32	11287.86
BIC	11282.84	11283.43	11949.69	11310.89	11946.34	11965.61	11311.14

```
Call:
lm(formula = Classical ~ as.factor(Harmony) + as.factor(Instrument) +
    as.factor(Voice), data = ratings2)

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)                 4.34016    0.12987  33.420  < 2e-16 ***
as.factor(Harmony)I-V-IV   -0.03108    0.13008  -0.239 0.811168
as.factor(Harmony)I-V-VI    0.76909    0.13008   5.913 3.83e-09 ***
as.factor(Harmony)IV-I-V    0.05007    0.12997   0.385 0.700092
as.factor(Instrument)piano  1.37359    0.11298  12.158  < 2e-16 ***
as.factor(Instrument)string 3.13312    0.11230  27.899  < 2e-16 ***
as.factor(Voice)par3rd     -0.41247    0.11271  -3.660 0.000258 ***
as.factor(Voice)par5th     -0.37058    0.11264  -3.290 0.001016 **
---
```

```
Signif. codes:  0 Ô***Õ 0.001 Ô**Õ 0.01 Ô*Õ 0.05 Ô.Õ 0.1 Ô Õ 1

Residual standard error: 2.297 on 2485 degrees of freedom
Multiple R-squared:  0.255,     Adjusted R-squared:  0.2529
F-statistic: 121.5 on 7 and 2485 DF,  p-value: < 2.2e-16
```

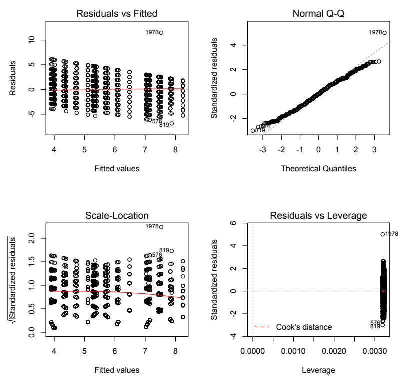

Figure 1: Diagnostic plots for the full model.

b.

i.

$Classical_i = \alpha_{j[i]} + \beta_1 * (Harmony_i) + \beta_2 * (Instrument_i) + \beta_3 * (Voice_i) + \epsilon$
$\epsilon \sim N(0, \sigma^2)$
$\alpha_j \sim \beta_0 + \lambda_j$
$\lambda_j \sim N(0, \tau^2)$

ii.

Below is the the summary output of adding in the random intercept into the model. Table 2 reports the AIC and BIC for both the non-random intercept model and random intercept model. We observe that both the AIC and BIC for the model including the random intercept is between 40 and 50 less than the model without the random intercept. We also observe the residuals in Figures 1 for the fixed effect model and Figures 2-4 for the random effects residual plots. We observe all three residual plots for the random effect model are relatively randomly scattered and centered at zero. We do not observe the trend in the residuals that we do with the fixed effect model. AIC, BIC, and the residual plots give us evidence to conclude the model with the random intercept is stronger than the model with fixed effects only.

Table 2: AIC and BIC of the Random and non-random intercept models

	No Random Intercept	Random Intercept
AIC	11230.45	11242.69
BIC	11282.84	11283.43

```
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) +      (1 | Subject)
   Data: ratings2

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)                   4.34374    0.18914   22.97
as.factor(Harmony)I-V-IV     -0.03251    0.10718   -0.30
as.factor(Harmony)I-V-VI      0.77096    0.10718    7.19
as.factor(Harmony)IV-I-V      0.04989    0.10709    0.47
as.factor(Instrument)piano    1.37705    0.09318   14.78
as.factor(Instrument)string   3.13161    0.09257   33.83
as.factor(Voice)par3rd       -0.41507    0.09287   -4.47
as.factor(Voice)par5th       -0.37439    0.09281   -4.03

Correlation of Fixed Effects:
                  (Intr) a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p as.fctr(Instrmnt)s
a.(H)I-V-IV       -0.282
a.(H)I-V-VI       -0.282  0.499
a.(H)IV-I-V       -0.283  0.499      0.499
as.fctr(Instrmnt)p -0.244  0.001      0.001     -0.001
as.fctr(Instrmnt)s -0.245 -0.001     -0.001     -0.001   0.498
as.fctr(V)3       -0.245 -0.002      0.001      0.002  -0.001             -0.001
as.fctr(V)5       -0.244 -0.002     -0.003     -0.001  -0.001              0.000
                  a.(V)3
a.(H)I-V-IV
a.(H)I-V-VI
a.(H)IV-I-V
as.fctr(Instrmnt)p
as.fctr(Instrmnt)s
as.fctr(V)3
as.fctr(V)5        0.500
```

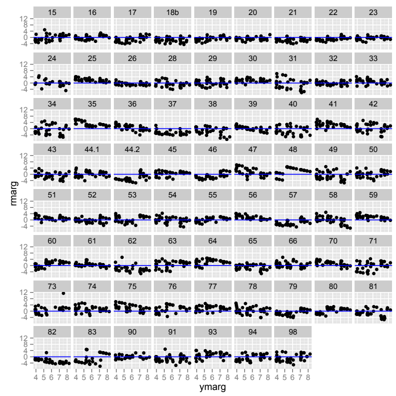

Figure 2: Marginal residuals for random intercept model.

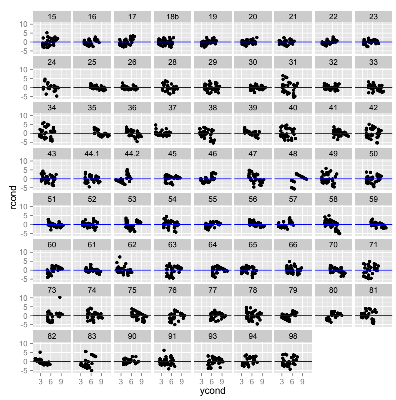

Figure 3: Conditional residuals for random intercept model.

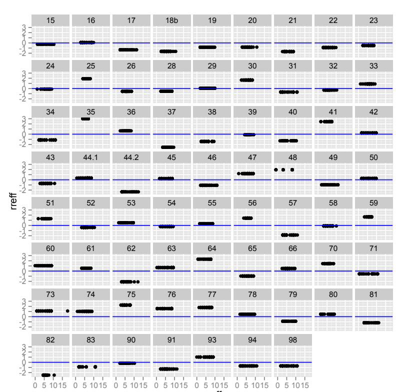

Figure 4: Random effect residuals versus fitted.

iii.

We start by comparing the full model with all three main experimental factors (instrument, harmony, and voice) on Classical ratings, using the repeated-measures model with the random intercept for participants. We then compare this model to every combination of the three variables by creating a total of 7 different models. All seven models are not shown but can be seen in the attached R appendix. The full summary output of the model we deemed most appropriate (the full model) is displayed below. Table 1 displays the AIC and BIC for each of the seven models.

When comparing AIC and BIC of the 7 models, we observe that the full model with harmony, instrument, and voice using the repeated-measures model with the random intercept for participants we observe both criterions prefer the full model. We move forward with our analysis with the full model.

Table 3: AIC and BIC of the 7 repeated-measures model with the random intercept for participant models

	Harm/Instr/Voice	Harm/Instr	Harm/Voice	Instr/Voice	Harm	Voice	Instr
AIC	10491.51	10505.58	11423.04	10552.74	11429.98	11461.42	10566.14
BIC	10549.73	10552.15	11469.6	10593.49	11464.91	11490.53	10595.25

```
> summary(lmer.1)
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) +      (1 | Subject)
   Data: ratings2

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)                   4.34374    0.18914   22.97
as.factor(Harmony)I-V-IV     -0.03251    0.10718   -0.30
as.factor(Harmony)I-V-VI      0.77096    0.10718    7.19
as.factor(Harmony)IV-I-V      0.04989    0.10709    0.47
as.factor(Instrument)piano    1.37705    0.09318   14.78
as.factor(Instrument)string   3.13161    0.09257   33.83
as.factor(Voice)par3rd       -0.41507    0.09287   -4.47
as.factor(Voice)par5th       -0.37439    0.09281   -4.03

Correlation of Fixed Effects:
                  (Intr) a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p as.fctr(Instrmnt)s a.(V)3
a.(H)I-V-IV       -0.282
a.(H)I-V-VI       -0.282  0.499
a.(H)IV-I-V       -0.283  0.499      0.499
as.fctr(Instrmnt)p -0.244  0.001      0.001     -0.001
as.fctr(Instrmnt)s -0.245 -0.001     -0.001     -0.001   0.498
as.fctr(V)3       -0.245 -0.002      0.001      0.002  -0.001             -0.001
as.fctr(V)5       -0.244 -0.002     -0.003     -0.001  -0.001              0.000              0.500
```

part c.
i.

We compare the model with a random intercept for each subject, the model with only fixed effects, and the model with three factors varying the intercept for each subject to account for the three personal biases in the experimental factors via AIC and BIC in Table 4. We also reported the summary output for the random intercept model varying for the three factors. We observe that the AIC and BIC for the random effect model with three factors varying the intercept for each subject is over a factor of 300 lower than the model with the random intercept only varying for each subject (not accounting for the three other factors). The full model is also stronger than the model with only fixed effects by a magnitude of over 400.

Table 4: AIC and BIC of the 7 repeated-measures model with the random intercept for participant models

	Random Intercept (Subject)	Only fixed effects	Random Intercept (Instrument, Harmony, Voice)
AIC	10491.51	11230.45	10075.51
BIC	10549.73	11282.84	10145.37

```
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) +      (1 | Subject:Instrument) +
(1 | Subject:Harmony) + (1 | Subject:Voice)  Data: ratings2

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)                4.34106    0.21435  20.252
as.factor(Harmony)I-V-IV  -0.03023    0.14317  -0.211
as.factor(Harmony)I-V-VI   0.77063    0.14316   5.383
as.factor(Harmony)IV-I-V   0.05618    0.14310   0.393
as.factor(Instrument)piano  1.36384    0.26232   5.199
as.factor(Instrument)string 3.12836    0.26203  11.939
as.factor(Voice)par3rd    -0.40699    0.08174  -4.979
as.factor(Voice)par5th    -0.37084    0.08168  -4.540

Correlation of Fixed Effects:
                  (Intr) a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p as.fctr(Instrmnt)s a.(V)3
a.(H)I-V-IV       -0.333
a.(H)I-V-VI       -0.333  0.499
a.(H)IV-I-V       -0.333  0.500      0.500
as.fctr(Instrmnt)p -0.611  0.000      0.000      0.000
as.fctr(Instrmnt)s -0.611  0.000      0.000      0.000   0.500
as.fctr(V)3       -0.190 -0.002      0.001      0.002  -0.001             0.000
as.fctr(V)5       -0.190 -0.001     -0.002     -0.001  -0.001             0.000              0.500
```

ii.

We start by comparing the full model with all three fixed main experimental factors (instrument, harmony, and voice) on Classical ratings, using the model with the random intercept for each subject to account for the three personal biases in the experimental factors for each model. We then compare this model to every combination of the three fixed effect variables by creating a total of 7 different models. All seven models are not shown but can be seen in the attached R appendix. The full summary output of the model we deemed most appropriate (the full model) is displayed below. Table 5 displays the AIC and BIC for each of the seven models.

When comparing AIC and BIC of the 7 models, we observe that the full model with all three fixed main experimental factors (instrument, harmony, and voice) on Classical ratings, using the model with the random intercept for each subject to account for the three personal biases in the experimental factors for each model has a higher AIC and BIC by a magnitude of 20 for AIC and 5 for BIC. We move forward with our analysis with the full model. The summary output for the model is shown in part 1c. part i. above.

Our interpretations for intercepts are how much of the random effect of harmony, instrument, and voice varies for each subject. We observe a variance of .44037 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the harmony. The random effect of harmony accounts for .44037 of the 2.43753 random effect variance for an average subject determining if a song is classical. We observe a variance of 0.02809 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the voice. The random effect of voice accounts for only 0.02809 of the 2.43753 random effect variance for an average subject determining if a song is classical. We observe a variance of 2.19850 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the instruments. The random effect of instruments accounts for 2.19850 of the 2.43753 random effect variance for an average subject determining if a song is classical. We observe instrument bias accounts for the majority of the random effect for each subject when predicting if a subject determines a song as classical. Overall, the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off their harmony, voice, and instrument bias.

Table 5: AIC and BIC of the 7 three factors varying he three fixed effect intercepts for harmony, instrument, and voice using the model with the random intercept for each subject to account for the three personal biases in the experimental factors for each model.

	Harm/Instr/Voice	Harm/Voice	Intr/Voice	Instr/Voice	Harm	Voice	Instr
AIC	10075.51	10092.66	10176.17	10101.74	10194.3	10204.66	10118.89
BIC	10145.37	10150.87	10234.38	10154.13	10240.87	10245.41	10159.64

iii.

$Classical_i = \alpha_{j[i]} + \beta_1 * (Harmony_i) + \beta_2 * (Instrument_i) + \beta_3 * (Voice_i) + \epsilon$
$\epsilon \sim N(0, \sigma^2)$
$\alpha_j \sim \beta_0 + \lambda_j + \psi_j + \phi_j$
$\lambda_j \sim N(0, \tau_1^2)$
$\psi_j \sim N(0, \tau_2^2)$
$\phi_j \sim N(0, \tau_3^2)$

Problem 2
a.

Before determining which individual covariates should be added to the model as fixed effects, we must investigate each covariate and determine if they are coded in a sensible for our model. Many of the covariates have additional missing values (beyond the original 27). The first variable we investigate is how proficient are you at your first instrument (X1stInstr). This variable has 1512 NA's but does not have any individuals coded as 0 (no instrument experience at all). Therefore we make the assumption of assigning all NA's to 0. We may want to explore if this is a smart decision with our investigator in future work.

We also code NA's to 0 for if you have taken an AP music class (0=no, 1 = yes), have you taken music classes in college (0=no, 1 = yes), How much did you concentrate on the instrument while listening (0-5, 0=not at all), and how much did you concentrate on the notes while listening (0-5, 0=not at all). Again, we may want to explore if this is a smart decision with our investigator in future work.

We combined and recoded the variables of have you heard Rob Paravonian's Pachelbel Rant and Have you heard Axis of Evil's comedy bit on the 4 Pachebel chords in popular music into one variable of if you heard of either of the two pieces (>1 for either then coded as 1), if you haven't heard of either (0 for both pieces, then coded as 0).

The variables of how familiar you are with Pachelbel's Canon in D and how much do you listen to classical music NA's were also coded as 0's or not at all. Again, we will want to check with our investigators. We then code the how much do you listen to classical music as either not at all (0, then coded as 0) or some (> 1 then coded as 1). We coded do you play piano, do you play guitar, and have you done any music composing in the same manor.

We do not use the difference between how much pop and rock from the 90s and 2000s do you listen to and the difference between that variable and how much 60s and 70s pop and rock you listen to in our model due to not knowing how to code the NA's in a reasonable manner. We also found the variables appeared to not have a great effect on predicting how classical the stimulus sounds in the presence of the other variables.
We then also take the score on a test of musical knowledge (OMSI) and determine that the variable needs transformed as the scale of OMSI is close to exponentially distributed. We determine with a log transformation then OMSI is close to equally distributed among it's range of values.

The first model we explore (which we will consider our full model) includes all fixed effect variables we believe may have an effect on our response. These variables include are you a musician, log(OMSI score), an interaction between are you a musician and log(OMSI score), have you taken ap music theory, how proficient are you at your first musical instrument, do you play the piano, do you play the guitar, have you composed music, and an auxiliary measure of luster's ability to distinguish classical vs. popular music. We compare the AIC and BIC of this full model along with the significance of the fixed effects in the presence of the other covariates while we eliminate one covariate at a time. We ultimately conclude that our final model includes do you play the guitar, do you play the piano, have you composed music, an auxiliary measure of luster's ability to distinguish classical vs. popular music, log(OMSI score), and are you a self-declared musician (along with the presence of the three factor variables harmony, voice and instrument as fixed effects and random intercepts). We compare the AIC and BIC of the full model, our final model, and model with know fixed effects (other than harmony, voice and instrument) in Table 6 below. Below is also the summary output of the final model.

We observe that the AIC and BIC are both lower for final model compared to the full model by a magnitude of 50 and 20 respectively. The final model has a smaller AIC then the basic model (by 8), but the basic model has a smaller BIC than the final model by 27. We expect the BIC to prefer simpler models, but we decide to still move forward with the final model because we determined that despite being more complex there are variables we feel the investigators would like to see present in the model, and these fixed effects were also significant (or close to it) in the final model. This is something we would want to talk about in great detail with our investigators and determine what variables they want to ultimately include in their final model.

Table 6: AIC and BIC of the 3 models with fixed and random intercepts for voice, harmony, and instrument for each subject.

	Full model	Basic model (no covariate fixed effects)	Final model (with selected covariate fixed effects)
AIC	10,082.5	10,075.51	10,067.63
BIC	10,222.21	10,145.37	10,172.41

```
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ pianoPlay + guitarPlay + Composing.noNA + X16.minus.17 +      log.omsi + Selfdeclare
+ as.factor(Harmony) + as.factor(Instrument) +      as.factor(Voice) + (1 | Subject:Instrument) + (1 | Subject:Harmony)
+      (1 | Subject:Voice)

REML criterion at convergence: 10031.63

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.43863  0.6623
 Subject:Voice      (Intercept) 0.02718  0.1649
 Subject:Instrument (Intercept) 1.91020  1.3821
 Residual                       2.43864  1.5616
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                           Estimate Std. Error t value
(Intercept)                 4.49875    0.51388   8.754
pianoPlay                   0.49009    0.26640   1.840
guitarPlay                  0.71175    0.31227   2.279
Composing.noNA              0.71089    0.27669   2.569
X16.minus.17               -0.10619    0.03708  -2.863
log.omsi                    0.14752    0.12039   1.225
Selfdeclare                -0.55175    0.13982  -3.946
as.factor(Harmony)I-V-IV   -0.03018    0.14274  -0.211
as.factor(Harmony)I-V-VI    0.77100    0.14273   5.402
as.factor(Harmony)IV-I-V    0.05611    0.14267   0.393
as.factor(Instrument)piano  1.36463    0.24612   5.545
as.factor(Instrument)string 3.12835    0.24582  12.726
as.factor(Voice)par3rd     -0.40725    0.08160  -4.991
as.factor(Voice)par5th     -0.37108    0.08153  -4.551

Correlation of Fixed Effects:
                  (Intr) pinPly gtrPly Cmp.NA X16..1 log.ms Slfdcl a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p
                  as.fctr(Instrmnt)s a.(V)3
pianoPlay          0.122
guitarPlay         0.075 -0.225
Compsng.nNA        0.169  0.058 -0.275
X16.mins.17       -0.066  0.015 -0.050 -0.054
log.omsi          -0.792 -0.113  0.035 -0.157  0.009
Selfdeclare       -0.014 -0.282 -0.239 -0.243 -0.078 -0.435
a.(H)I-V-IV       -0.138  0.000  0.000  0.001  0.001  0.000 -0.001
a.(H)I-V-VI       -0.139  0.000  0.000  0.001  0.001  0.000 -0.001  0.499
a.(H)IV-I-V       -0.139  0.000  0.000  0.000  0.001  0.000  0.000  0.500      0.500
as.fctr(Instrmnt)p -0.240 -0.001 -0.001  0.001 -0.001  0.001  0.000  0.000      0.000      0.000
```

as.fctr(Instrmnt)s	−0.240	0.000	0.000	−0.001	−0.001	0.000	0.001	0.000	0.000	0.000			
0.500													
as.fctr(V)3	−0.079	0.000	−0.001	0.001	0.001	0.000	0.000	−0.002	0.001	0.002	−0.001		
0.000													
as.fctr(V)5	−0.079	0.000	0.000	0.000	0.001	0.000	0.000	−0.001	−0.002	−0.001	−0.001		
0.000	0.500												

b.

We compare the 7 repeated-measures model with the random intercept for participant models and random intercept model with only subject. Only the full model with the three random intercepts and the model with only subject as the random intercept are shown in Table 7 below. We observe that the AIC and BIC are lower by over 400 for both the random intercept model with instrument, harmony and voice compared to the random intercept for only subject. We conclude the full model with the three random effects is the optimal model to use going forward.

Table 7: AIC and BIC of the 7 repeated-measures model with the random intercept for participant models and random intercept model with only subject. Only the full model with the three random intercepts and the model with only subject as the random intercept are shown below.

	Random Intercept (Instrument, Harmony, Voice)	Random Intercept (Only Subject)
AIC	10067.63	10494.68
BIC	10172.41	10587.82

c.

Using the final model output produced in 2a above we interpret the final coefficients. First we will interpret the fixed effects. Someone that has played the piano player, has on average, .49 higher classical rating then a person that does not play the guitar, controlling for the other covariates of the model. Someone that has played the guitar player, has on average, .711 higher classical rating then a person that does not play the guitar, controlling for the other covariates of the model. Someone that has composed music, has on average, .71 higher classical rating then a person that does not play the guitar, controlling for the other covariates of the model. For a one unit increase in Auxiliary measure of listener's ability to distinguish classical vs popular music, the classical rating decreases on average by 0.11, when controlling for all other variables. For a one unit increase in log(OMSI) to distinguish classical vs popular music, the classical rating increases on average by 0.15, when controlling for all other variables. For a one unit increase in the self-declared musician rating, we expect the classical rating to decrease by 0.55 on average, when controlling for all other variables. Being exposed to a harmony of I-V-IV decreases the classical rating by .03 compared to if the exposure is I-VI-V on average, when controlling for all other variables. Being exposed to a harmony of I-V-VI increases the classical rating by .77 compared to if the exposure is I-VI-V on average, when controlling for all other variables. Being exposed to a harmony of I-V-IV increases the classical rating by .06 compared to if the exposure is I-VI-V on average, when controlling for all other variables. Being exposed to a piano increase the classical rating by 1.36 compared to if they were exposed to an electric guitar on average, when controlling for all other variables. Being exposed to a string guitar increase the classical rating by 3.12 compared to if they were exposed to a electric guitar on average, when controlling for all other variables. Being exposed to a parallel 3rd voice decreases the classical rating by .41 compared to if they were exposed to a contrary motion on average, when controlling for all other variables. Being exposed to a parallel 5ths voice decreases the classical rating by .37 compared to if they were exposed to a contrary motion on average, when controlling for all other variables.

Next we will interpret the random effects. We observe a variance of 0.44 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the harmony. The random effect of harmony accounts for 0.44 of the 2.44 random effect variance for an average subject determining if a song is classical. We observe a variance of 0.027 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the voice. The random effect of voice accounts for only 0.027 of the 2.44 random effect variance for an average subject determining if a song is classical. We observe a variance of 1.91 for the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off the instruments. The random effect of instruments accounts for 1.91 of the 2.44 random effect variance for an average subject determining if a song is classical. We observe instrument bias accounts for the majority of the random effect for each subject when predicting if a subject determines a song as classical. Overall, the "average" intercept for a subject which gives us an idea how much our intercept is shifted up or down for each subject determining if a song is classical based off their harmony, voice, and instrument bias.

Problem 3

We dichotomize “Self-declare" ("are you a musician?") so that about half the participants are categorized as self-declared musicians and half are not by assigning everyone that are coded as not at all a musician (0), then assign them to 0 in the dichotomized variable and if they rated themselves as a musician at all (1 or greater out of 5) then assign them a 1. This splits self-declare in roughly two equal groups. We then explore any interactions between the dichotomized musician variable and other predictors in the model by trying each interaction combination. Our final model chosen includes a significant interaction between the self-declare variable and the auxiliary measure of listener's ability to distinguish classical versus popular music. Below we show the output of the final model and table 8 compares the AIC and BIC with and without the interaction term included.

We observe that the AIC of the model with the interaction decreases by a factor of 10 in comparison to the same model without the interaction term but the BIC increases by 7. The summary output below also shows us that the interaction effect is significant (t-value = -3.994) controlling for the other effects in the model, and thus we ultimately decide to include the interaction term in our final model. The interaction tells us people that don't declare themselves as musicians but have high auxiliary music scores, on average tend to believe the stimuli sound more classical compared to if a person declares themselves a musician and the higher they do on the auxiliary test (a strong musician), on average they tend to believe the stimuli sound less classical.

Table 8: AIC and BIC of the model with an interaction between the dichotomized musician variable and the auxiliary measure of listener's ability to distinguish classical versus popular music.

	No Interaction	Interaction
AIC	10079.01	10068.74
BIC	10172.41	10179.34

```
> summary(lmer.slope.final.SelfDeclare.int)
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ pianoPlay + guitarPlay + Composing.noNA + Selfdeclare2 *      X16.minus.17 + log.omsi + Selfdeclare
as.factor(Instrument) + as.factor(Voice) + (1 | Subject:Instrument) +      (1 | Subject:Harmony) + (1 | Subject:Voice)

REML criterion at convergence: 10030.74

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.43834  0.6621
 Subject:Voice      (Intercept) 0.02729  0.1652
 Subject:Instrument (Intercept) 1.89466  1.3765
 Residual                       2.43859  1.5616
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                            Estimate Std. Error t value
(Intercept)                 4.065384   0.546132   7.444
pianoPlay                   0.297111   0.258059   1.151
guitarPlay                  0.669162   0.315631   2.120
Composing.noNA              0.581404   0.272786   2.131
Selfdeclare2                0.088026   0.345433   0.255
X16.minus.17               -0.005229   0.045538  -0.115
log.omsi                    0.009066   0.119516   0.076
as.factor(Harmony)I-V-IV   -0.030043   0.142714  -0.211
as.factor(Harmony)I-V-VI    0.770993   0.142698   5.403
as.factor(Harmony)IV-I-V    0.056206   0.142641   0.394
as.factor(Instrument)piano  1.365649   0.245218   5.569
as.factor(Instrument)string 3.127994   0.244914  12.772
as.factor(Voice)par3rd     -0.407112   0.081616  -4.988
as.factor(Voice)par5th     -0.371168   0.081552  -4.551
```

```
Selfdeclare2:X16.minus.17    -0.317923    0.079598   -3.994

Correlation of Fixed Effects:
                    (Intr) pinPly gtrPly Cmp.NA Slfdc2 X16..1 log.ms a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p as
pianoPlay            0.059
guitarPlay          -0.032 -0.251
Compsng.nNA          0.090  0.020 -0.276
Selfdeclar2          0.289 -0.123 -0.207 -0.149
X16.mins.17         -0.122  0.036  0.032 -0.007  0.132
log.omsi            -0.897 -0.179  0.045 -0.188 -0.397  0.017
a.(H)I-V-IV         -0.130  0.000  0.000  0.001  0.000  0.002  0.000
a.(H)I-V-VI         -0.131  0.000  0.000  0.001  0.000  0.001  0.000  0.499
a.(H)IV-I-V         -0.130  0.000  0.000  0.000  0.000  0.001  0.000  0.500       0.500
as.fctr(Instrmnt)p  -0.225 -0.001 -0.001  0.001  0.002  0.000  0.000  0.000       0.000      0.000
as.fctr(Instrmnt)s  -0.224  0.000  0.000 -0.001  0.000 -0.002  0.000  0.000       0.000      0.000
0.500
as.fctr(V)3         -0.075  0.000 -0.001  0.001  0.000  0.001  0.000 -0.002       0.001      0.002   -0.001
0.000
as.fctr(V)5         -0.074  0.000  0.000  0.000 -0.001  0.001  0.000 -0.001      -0.002     -0.001   -0.001
0.000               0.500
Sl2:X16..17          0.058 -0.044 -0.099 -0.055 -0.404 -0.577  0.016 -0.001       0.000      0.000   -0.001
0.001              -0.001
                    a.(V)5
pianoPlay
guitarPlay
Compsng.nNA
Selfdeclar2
X16.mins.17
log.omsi
a.(H)I-V-IV
a.(H)I-V-VI
a.(H)IV-I-V
as.fctr(Instrmnt)p
as.fctr(Instrmnt)s
as.fctr(V)3
as.fctr(V)5
Sl2:X16..17          0.000
```

Problem 4

a

Now we re-examine the data in term of the “Popular" ratings, instead of the “Classical" ratings, using similar hierarchical linear models. We now look at the influence of instrument, harmony, and voice on popular ratings by comparing AIC and BIC below of each model in table 9. The models we compare all have the three random effects for instrument, harmony, and voice but explore the various combinations of instrument, harmony, and voice as fixed effects (7 models without other covariate fixed effects). According to AIC and BIC, the model with only instrument is the best model, but we ultimately decide to use the full model because the three variables are the stimuli of interest in the study and it would not make sense to remove them from our analysis. The output for the full model can also be found below.

Table 9: AIC and BIC of the 7 three factors varying he three fixed effect intercepts for harmony, instrument, and voice using the model with the random intercept for each subject to account for the three personal biases in the experimental factors for each model.

	Harm/Instr/Voice	Harm/Voice	Intr/Voice	Instr/Voice	Harm	Voice	Instr
AIC	10,097.24	10,091.75	10,177.79	10,089.39	10,172.41	10,170.06	10,083.91
BIC	10,167.09	10,149.96	10,236	10,141.78	10,218.98	10,210.81	10,124.66

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

REML criterion at convergence: 10073.24

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41144  0.6414
 Subject:Voice      (Intercept) 0.03226  0.1796
 Subject:Instrument (Intercept) 1.99988  1.4142
 Residual                       2.49033  1.5781
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                6.57991    0.20709   31.77
as.factor(Harmony)I-V-IV  -0.02557    0.14059   -0.18
as.factor(Harmony)I-V-VI  -0.27156    0.14057   -1.93
as.factor(Harmony)IV-I-V  -0.18545    0.14051   -1.32
as.factor(Instrument)piano -0.94900   0.25153   -3.77
as.factor(Instrument)string -2.60587  0.25122  -10.37
as.factor(Voice)par3rd     0.16380    0.08324    1.97
as.factor(Voice)par5th     0.16206    0.08317    1.95

Correlation of Fixed Effects:
                  (Intr) a.(H)I-V-I a.(H)I-V-V a.(H)IV as.fctr(Instrmnt)p as.fctr(Instrmnt)s a.(V)3
a.(H)I-V-IV       -0.338
a.(H)I-V-VI       -0.339  0.499
a.(H)IV-I-V       -0.339  0.500      0.500
as.fctr(Instrmnt)p -0.606  0.000     0.000      0.000
as.fctr(Instrmnt)s -0.607  0.000     0.000      0.000   0.500
as.fctr(V)3       -0.200 -0.002      0.001      0.002  -0.001             0.000
as.fctr(V)5       -0.200 -0.001     -0.002     -0.001  -0.001             0.000              0.500
```

what does this tell you about the influence

b.

In table 10 we compare the AIC and BIC of the full model from part a above, the optimal model in terms of AIC/BIC above (only interment fixed effect), and all the full model above plus the fixed effects after optimizing model selection in terms of AIC and BIC. We determined to leave the fixed effects for log(OMSI), the auxiliary score, and self-declare because they were important covariates in the context of the study, and we also included the combined variables of have you heard Rob Paravonian's Pachelbel Rant and Have you heard Axis of Evil's comedy bit on the 4 Pachebel chords in popular music because it was significant. All with the three random intercepts for Harmony, Instrument, and Voice. The we ultimately found that the added fixed effects make the model worse than the model without any additional fixed covariates. In additional to using our final model in part 2c, we also did variable selection again and found that no model with fixed effects out-performed the two models below without any additional fixed covariates.

Table 10: AIC and BIC of the full model from part a above, the optimal model in terms of AIC/BIC above (only interment fixed effect), and all the full model above plus the fixed effects from the final model in part 2c. All with the three random intercepts for Harmony, Instrument, and Voice

interpretation needed

	Harm/Instr/Voice (w/ other fixed covariates)	Instr (w/ other fixed covariates)	Harm/Instr/Voice (w/ other fixed covariates)
AIC	10,097.24	10,083.91	10099.63
BIC	10,167.09	10,124.66	10192.77

c.

Once again we dichotomize "Self-declare" ("are you a musician?") so that about half the participants are categorized as self-declared musicians and half are not by assigning everyone that are coded as not at all a musician (0), then assign them to 0 in the dichotomized variable and if they rated themselves as a musician at all (1 or greater out of 5) then assign them a 1. This splits self-declare in roughly two equal groups. We then explore any interactions between the dichotomized musician variable and other predictors in the model by trying each interaction combination. Our final model chosen includes a significant interaction between the self-declare variable and the auxiliary measure of listener's ability to distinguish classical versus popular music.

Below we show the model with all of the fixed covariate effects that we had in our model in part 4b above, and we determined that there is an interaction effect between self-declare and if they have heard Rob Paravonian's Pachelbel Rant or heard Axis of Evil's comedy bit on the 4 Pachebel chords in popular music. We found a significant interaction effect and the AIC of this model is slightly lower (2) than the our full model before. It also has the same AIC as the model with only harmony, instrument, and voice as fixed effects but does worse in terms of BIC (as we expect with additional covariates in the model). We ultimately conclude that this model with the interaction effect to be our final model for predicting how popular does the stimulus sound due to the importance of the covariates be included in our study. We will want to check with our investigator to determine if our decision of the importance of these additional covariates is a valid decision. The final model output is also below.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Popular ~ Selfdeclare2 * KnowRobOrAxis + X16.minus.17 + log.omsi +       Selfdeclare2 + as.factor(Harmony)
+ as.factor(Instrument) +       as.factor(Voice) + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
(1 | Subject:Voice)

REML criterion at convergence: 10063.18

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.40942  0.6399
 Subject:Voice      (Intercept) 0.03175  0.1782
```

This looks reasonable but it would be useful to know how you arrived at this model.

Table 11: AIC and BIC of the full model from part a above, the optimal model in terms of AIC/BIC above (only interment fixed effect), and all the full model above plus the fixed effects from the final model in part 2c. All with the three random intercepts for Harmony, Instrument, and Voice

	Harm/Instr/Voice (w/ other fixed covariates)	Instr (w/ other fixed covariates)	Harm/Instr/Voice (w/ interaction)
AIC	10,097.24	10,083.91	10097.18
BIC	10,167.09	10,124.66	10196.14

```
 Subject:Instrument (Intercept) 1.85413  1.3617
 Residual                       2.49091  1.5783
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                              Estimate Std. Error t value
(Intercept)                    5.85190    0.53828  10.872
Selfdeclare2                  -0.08980    0.31559  -0.285
KnowRobOrAxis                 -0.01876    0.38351  -0.049
X16.minus.17                   0.05609    0.03768   1.488
log.omsi                       0.10285    0.11570   0.889
as.factor(Harmony)I-V-IV      -0.02512    0.14039  -0.179
as.factor(Harmony)I-V-VI      -0.27125    0.14037  -1.932
as.factor(Harmony)IV-I-V      -0.18538    0.14031  -1.321
as.factor(Instrument)piano    -0.94838    0.24311  -3.901
as.factor(Instrument)string   -2.60612    0.24280 -10.734
as.factor(Voice)par3rd         0.16394    0.08316   1.971
as.factor(Voice)par5th         0.16223    0.08309   1.952
Selfdeclare2:KnowRobOrAxis     1.02366    0.51488   1.988

Correlation of Fixed Effects:
                   (Intr) Slfdc2 KnwROA X16..1 log.ms a.(H)I-V-I a.(H)I-V-V a.(H)IV
                   as.fctr(Instrmnt)p as.fctr(Instrmnt)s a.(V)3 a.(V)5
Selfdeclar2         0.332
KnowRbOrAxs        -0.011  0.242
X16.mins.17        -0.110 -0.079  0.220
log.omsi           -0.899 -0.510 -0.138 -0.002
a.(H)I-V-IV        -0.130  0.000  0.000  0.001  0.000
a.(H)I-V-VI        -0.130  0.000  0.000  0.001  0.000  0.499
a.(H)IV-I-V        -0.130  0.000  0.000  0.001  0.000  0.500      0.500
as.fctr(Instrmnt)p -0.226  0.001  0.000 -0.001  0.000  0.000      0.000      0.000
as.fctr(Instrmnt)s -0.226  0.000  0.000 -0.001  0.000  0.000      0.000      0.000   0.500
as.fctr(V)3        -0.077  0.000  0.000  0.001  0.000 -0.002      0.001      0.002  -0.001
0.000
as.fctr(V)5        -0.077 -0.001  0.000  0.001  0.000 -0.001     -0.002     -0.001  -0.001
0.000              0.500
Slfdc2:KROA         0.070 -0.453 -0.744 -0.199  0.039  0.000      0.000      0.000   0.000
0.000              0.000  0.000
```

Problem 5

Through our analysis we investigated how various musical stimuli, along with other covariates effected a recruited sample of undergraduates at the University of Pittsburgh's ability to determine how classical the music sounds and how popular the music sounds. We do this by building two models, one for predicting how classical the student's believed the music sounds and one for predicting how popular the student's believed the music sounds. Before the study Dr. Jimenez had multiple hypotheses of interest. These included that instrument should have the largest influence on rating. The harmonic progression I-V-iv may be frequently rated as classical as it is the beginning progression of Pachelbel's Canon in D. Although it is also in many popular songs. Dr. Jimenez also believes the voice leading category, contrary motion, will be frequently rated as classical. We keep these hypotheses in mind while conducting our analysis.

In both of our models for predicting if a student believes a music sound is classical or popular, we determined that we should include additional variance components for harmony, voice, and instrument because this will attempt to account for "personal biases" in ratings (i.e. perhaps person A is more included to rate everything as classical, and person B is more included to rate everything as popular). We also explored other covariates that are important in predicting how likely someone is to determine a sound is classical or popular. In our classical model we found if a student played the guitar, played the piano, or composed music, then it leads to them saying musical sounds more classical, on average (each relationship holding the rest of the covariates constant). Although the higher the auxiliary measure of the listener's ability to distinguish classical vs popular music, the lower they will rate the sound classical. We also observe an interaction between if people considered themselves musicians and their auxiliary music score. The interaction tells us given someone has a high auxiliary score, they give a lower classical rating then someone with a high auxiliary score but doesn't consider themselves a musician.

In terms of the stimuli, we observed that the Canon in D harmony I-V-IV lead to a higher classical rating on average then the baseline harmonic measure of I-VI-V, confirming Dr. Jimenez's hypothesis. We also observe that if the stimuli is a string quartet then the music is rated to sound 3 rating points higher than a music sound with an electric guitar. This is the highest effect on classical rating (relative to baseline) compared to the other stimuli factors and also confirms Dr. Jimenez's hypothesis. Lastly, we found that the other two voice lead stimuli lead to a lower classical rating score than contrary motion which also confirms Dr. Jimenez's hypothesis. It is also important to point out that the random effect for instrument lead to the most variance in the random intercept of our model, or the type of instrument lead to the most bias among subjects (i.e. people vary in the degree in which they are inclined to call music played by a string quartet "classical").

When building our model for predicting the popular score rating we observe the opposite relationship among the stimuli as our classical model above (as expected), and once again Dr. Jimenez's predictions are confirmed. Although we do determine slightly different covariates determined the popular ratings score. We include the logged OMSI score and the auxiliary score in our final model because we assume these covariates would be of importance to Dr. Jiminez, although their effects are not significant. Although we do find a significant interaction between if a student called themselves a musician and if they have heard of either of the comedic acts. This can be interpreted as if a person calls themselves a musician but has not heard of the comedic examples or they have heard of the comedic examples but are not a musician then they give a stimuli a lower popular score on average then someone that considers themselves a musician and has seen the comedic acts.

Overall we have built two statistical that investigate how various musical stimuli, along with other covariates effected a recruited sample of undergraduates at the University of Pittsburgh's ability to determine how classical the music sounds and how popular the music sounds. We were able to confirm Dr. Jimenez's hypotheses and determine what other covariates help predict how classical or popular a student believes a stimuli sounds.

4: 18
5: 20

38

be careful in your writeup in part 5 to make sure grammar, usage, spelling etc, are all correct.

```
# Mike Pane
# 12/5/13
# Final Exam

ratings <- read.csv("ratings.csv", header=TRUE)


ratings2 <- ratings[which(!is.na(ratings$Classical)),]

#1
out <- ratings2[1978,]

lm.1 <- lm(Classical~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice), data=ratings2)
lm.2.reduced <- lm(Classical ~ as.factor(Harmony) + as.factor(Instrument), data=ratings2)
lm.3.reduced <- lm(Classical ~ as.factor(Harmony) + as.factor(Voice), data=ratings2)
lm.4.reduced <- lm(Classical ~ as.factor(Instrument) + as.factor(Voice), data=ratings2)
lm.5.reduced <- lm(Classical ~ as.factor(Harmony), data=ratings2)
lm.6.reduced <- lm(Classical ~ as.factor(Voice), data=ratings2)
lm.7.reduced <- lm(Classical ~ as.factor(Instrument), data=ratings2)

list.models <- list(lm.1,lm.2.reduced, lm.3.reduced, lm.4.reduced,lm.5.reduced, lm.6.reduced, lm.7.reduced)
lapply(list.models, AIC)
lapply(list.models, BIC)

par(mfrow=c(2,2))
plot(lm.1)

# R^2 adjusted full model is better
# Compare AIC and BIC. Both prefer the full model. Comment on outlier being infuential
# Do I want to make boxplots?
# Check residuals

#b

library(arm)
lmer.1 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) + (1|Subject), data=ratings2)
summary(lmer.1)

AIC(lmer.1)
AIC(lm.1)
BIC(lmer.1)
BIC(lm.1)

r.marg <- function(m) {
  y <- m@frame[,1]
  yhat <- model.matrix(m) %*% fixef(m)
  return(y-yhat)
}

r.cond <- function(m) {residuals(m)}

r.reff <- function(m) {r.marg(m) - r.cond(m)}

# suitable fitted values to plot them against...
# (you can plot them against other things as well...

yhat.marg <- function(m) { model.matrix(m) %*% fixef(m) }

yhat.cond <- function(m) {
  y <- m@frame[,1]
  y - r.cond(m)
}

yhat.reff <- function(m) { yhat.marg(m) + r.cond(m) }

##################
par(mfrow=c(1,3))
mod= lmer.1

library(ggplot2)

plotdata <- data.frame(mod@frame, fixed.re = fitted(lmer.1),rmarg = r.marg(mod), rcond = r.cond(mod), rreff = r.reff(mo

qplot(data = plotdata, x = ymarg, y = rmarg, facets = ~ Subject) + geom_abline(slope=0, intercept=0, colour="blue")
qplot(data = plotdata, x = ycond, y = rcond, facets = ~ Subject)+ geom_abline(slope=0, intercept=0, colour="blue")
qplot(data = plotdata, x = yreff, y = rreff, facets = ~ Subject)+ geom_abline(slope=0, intercept=0, colour="blue")

xyplot(r.reff(lmer.1)~yhat.reff(lmer.1)|as.factor(ratings2$Subject))

# lm
par(mfrow=c(2,2))
plot(lm.1)


#######

# iii.
lmer.1 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) + (1|Subject), data=ratings2)
lmer.2 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument)  + (1|Subject), data=ratings2)
lmer.3 <- lmer(Classical ~ as.factor(Harmony)  + as.factor(Voice) + (1|Subject), data=ratings2)
lmer.4 <- lmer(Classical ~ as.factor(Instrument) + as.factor(Voice) + (1|Subject), data=ratings2)
lmer.5 <- lmer(Classical ~ as.factor(Harmony)  + (1|Subject), data=ratings2)
```

```
lmer.6 <- lmer(Classical ~  as.factor(Voice) + (1|Subject), data=ratings2)
lmer.7 <- lmer(Classical ~  as.factor(Instrument)  + (1|Subject), data=ratings2)

list.models.lmer <- list(lmer.1,lmer.2,lmer.3,lmer.4,lmer.5,lmer.6,lmer.7)
lapply(list.models.lmer, AIC)
lapply(list.models.lmer, BIC)

summary(lmer.1)

# Full model is the best using both AIC and BIC!

###########
# Part c. #
###########

#i.
lmer.slope.1 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) + (1|Subject:Instrument)


list.models.lmer.all <- list(lmer.1,lm.1,lmer.slope.1)
lapply(list.models.lmer.all, AIC)
lapply(list.models.lmer.all, BIC)

summary(lmer.slope.1)

# Lmer with multiple random slopes for each category does better.

# ii
lmer.slope.2 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument) + (1|Subject:Instrument) + (1|Subject:Harmo
lmer.slope.3 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) +
lmer.slope.4 <- lmer(Classical ~ as.factor(Instrument) + as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony
lmer.slope.5 <- lmer(Classical ~ as.factor(Harmony) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice),
lmer.slope.6 <- lmer(Classical ~ as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), d
lmer.slope.7 <- lmer(Classical ~ as.factor(Instrument) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voic

list.models.lmer.all <- list(lmer.slope.1, lmer.slope.2,lmer.slope.3,lmer.slope.4,lmer.slope.5,lmer.slope.6,lmer.slope.
lapply(list.models.lmer.all, AIC)
lapply(list.models.lmer.all, BIC)

# Full model is best!

##################################

# Problem 2

#a

# reduced
lmer.slope.1 <- lmer(Classical ~ as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) + (1|Subject:Instrument)

#Full
# Assess missing values.
table(ratings2$X1stInstr)
# The coded variable can have 0 (no intstrument experience) but the data does not have any coded individuals as 0.
# They have 1512 NA's
# in order to keep sample size consistent when comparing models, we assign NA's to 0.
# This makes sense because we don't have any original coded 0s.

X1stIntr.noNA <- ratings2$X1stInstr
X1stIntr.noNA[is.na(ratings2$X1stInstr)==TRUE] <- 0

APTheory.noNA <- ratings2$APTheory
APTheory.noNA[is.na(ratings2$APTheory)==TRUE] <- 0

NoClass.noNA <- ratings2$NoClass
NoClass.noNA[is.na(ratings2$NoClass)==TRUE] <- 0
NoClass.noNA[NoClass.noNA>0] <- 1

CollegeMusic.noNA <- ratings2$CollegeMusic
CollegeMusic.noNA[is.na(ratings2$CollegeMusic)==TRUE] <- 0

# We are not including 1990s2000s because of NA's

# Explain coding here.
know.axis2 <- ifelse (ratings2$KnowAxis < 1, 0, 1)
KnowAxis.noNA <- know.axis2
KnowAxis.noNA[is.na(know.axis2)==TRUE] <- 0

know.rob2 <- ifelse (ratings2$KnowRob < 1, 0, 1)
KnowRob.noNA <- know.rob2
KnowRob.noNA[is.na(know.rob2)==TRUE] <- 0

composing2 <- ifelse(ratings2$Composing <= 0, 0, 1)
Composing.noNA <- composing2
Composing.noNA[is.na(composing2)==TRUE] <- 0

# Instr.minus.notes was calculated by subbing in 0's for ConsInstr and ConsNotes for missing values.
# We will take same stratgegy

table((ratings2$ConsInstr - ratings2$ConsNotes) == ratings2$Instr.minus.Notes)

ConsInstr.naNA <- ratings2$ConsInstr
```

```
ConsInstr.naNA[is.na(ratings2$ConsInstr)==TRUE] <- 0

ConsNotes.naNA <- ratings2$ConsInstr
ConsNotes.naNA[is.na(ratings2$ConsNotes)==TRUE] <- 0

clsListen2 <- ratings2$ClsListen
clsListen2[clsListen2 > 1] <- 2
clsListen.noNA <- clsListen2
clsListen.noNA[is.na(clsListen2==TRUE)] <- 0


# Rest don't have NA's

# Remaining variable coding for non-NA variables
# TODO: Explain reasoning

log.omsi <- log(ratings2$OMSI)
pianoPlay <- ifelse(ratings2$PianoPlay < 1, 0, 1) # No NA's
guitarPlay <- ifelse(ratings2$GuitarPlay < 1, 0, 1) # No NA's
clsListen2 <- ifelse(ratings2$ClsListen < 1, 0, 1)

#Combine KnowRob and KnowAxis
# Do they know either of them?
KnowRobOrAxis <- rep(0,length(KnowRob.noNA))
KnowRobOrAxis[KnowRob.noNA == 1 | KnowAxis.noNA == 1 ] <- 1

attach(ratings2)
lmer.slope.fixEF.1 <- lmer(Classical ~ Selfdeclare*log.omsi +  X1stIntr.noNA + APTheory.noNA + KnowRobOrAxis
+
                             pianoPlay + guitarPlay + Composing.noNA +  as.factor(NoClass.noNA)  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))
summary(lmer.slope.fixEF.1)
AIC(lmer.slope.fixEF.1)
BIC(lmer.slope.fixEF.1)

lmer.slope.fixEF.2 <- lmer(Classical ~ X1stIntr.noNA + APTheory.noNA + KnowRobOrAxis  +
                             pianoPlay + guitarPlay + Composing.noNA +  as.factor(NoClass.noNA)  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))

summary(lmer.slope.fixEF.2)
AIC(lmer.slope.fixEF.2) # 10086.5
BIC(lmer.slope.fixEF.2) # 10220.38


lmer.slope.fixEF.3 <- lmer(Classical ~ X1stIntr.noNA  + KnowRobOrAxis  +
                             pianoPlay + guitarPlay + Composing.noNA +  as.factor(NoClass.noNA)  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))

summary(lmer.slope.fixEF.3)
AIC(lmer.slope.fixEF.3)
# 10084.19
BIC(lmer.slope.fixEF.3)
# 10212.26

#####
lmer.slope.fixEF.4 <- lmer(Classical ~ X1stIntr.noNA  + KnowRobOrAxis  +
                             pianoPlay + guitarPlay + Composing.noNA +  as.factor(NoClass.noNA)  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))

summary(lmer.slope.fixEF.4)
AIC(lmer.slope.fixEF.4)
# 10083.95
BIC(lmer.slope.fixEF.4)
# 10212.02

###
lmer.slope.fixEF.5 <- lmer(Classical ~ X1stIntr.noNA  +
                             pianoPlay + guitarPlay + Composing.noNA +  as.factor(NoClass.noNA)  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))

summary(lmer.slope.fixEF.5)
AIC(lmer.slope.fixEF.5)
# 10082.24
BIC(lmer.slope.fixEF.5)
# 10204.48

########
lmer.slope.fixEF.6 <- lmer(Classical ~ X1stIntr.noNA  +
                             pianoPlay + guitarPlay + Composing.noNA  + clsListen.noNA +
                             X16.minus.17 + log.omsi + Selfdeclare  + as.factor(Harmony) + as.factor(Instrument) +
                             as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                           + (1|Subject:Voice))
```

```
summary(lmer.slope.fixEF.6)
AIC(lmer.slope.fixEF.6)
# 10072.61
BIC(lmer.slope.fixEF.6)
# 10189.04

######
lmer.slope.fixEF.7 <- lmer(Classical ~ X1stIntr.noNA +
                              pianoPlay + guitarPlay + Composing.noNA +
                              X16.minus.17 + log.omsi + Selfdeclare + as.factor(Harmony) + as.factor(Instrument) +
                              as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                            + (1|Subject:Voice))
summary(lmer.slope.fixEF.7)
AIC(lmer.slope.fixEF.7)
# 10070.43
BIC(lmer.slope.fixEF.7)
# 10181.03

#####

lmer.slope.fixEF.8 <- lmer(Classical ~
                              pianoPlay + guitarPlay + Composing.noNA +
                              X16.minus.17 + log.omsi + Selfdeclare + as.factor(Harmony) + as.factor(Instrument) +
                              as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                            + (1|Subject:Voice))
summary(lmer.slope.fixEF.8)
AIC(lmer.slope.fixEF.8)
# 10067.63
BIC(lmer.slope.fixEF.8)
# 10172.41

############

#b
lmer.slope.final.1 <- lmer(Classical ~
                              pianoPlay + guitarPlay + Composing.noNA +
                              X16.minus.17 + log.omsi + Selfdeclare + as.factor(Harmony) + as.factor(Instrument) +
                              as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                            + (1|Subject:Voice))

summary(lmer.slope.final.1)
AIC(lmer.slope.final.1)
# 10067.63
BIC(lmer.slope.final.1)
# 10172.41

####################

# not this model!
lmer.slope.final.2 <- lmer(Classical ~
                              pianoPlay + guitarPlay + Composing.noNA +
                              X16.minus.17 + log.omsi + Selfdeclare + as.factor(Harmony) + as.factor(Instrument) +
                              as.factor(Voice) + (1|Subject))

summary(lmer.slope.final.2)
AIC(lmer.slope.final.2)
# 10494.68
BIC(lmer.slope.final.2)
# 10587.82

################################################################
# Problem 3 -- With self declare
 Selfdeclare2 <- rep(0, length(Selfdeclare))
 Selfdeclare2[Selfdeclare > 2] <- 1

lmer.slope.final.SelfDeclare <- lmer(Classical ~ pianoPlay + guitarPlay + Composing.noNA + X16.minus.17 + log.omsi + S
+ as.factor(Harmony) + as.factor(Instrument) +as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subj

summary(lmer.slope.final.SelfDeclare)
AIC(lmer.slope.final.SelfDeclare)
# 10079.01
BIC(lmer.slope.final.SelfDeclare)
# 10172.41

### Interactions?
lmer.slope.final.SelfDeclare.int <- lmer(Classical ~
                              pianoPlay + guitarPlay + Composing.noNA +
                              Selfdeclare2*X16.minus.17 + log.omsi + Selfdeclare2 + as.factor(Harmony) + as.factor(Instr
                              as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony)
                            + (1|Subject:Voice))

summary(lmer.slope.final.SelfDeclare.int)
AIC(lmer.slope.final.SelfDeclare.int)
# 10068.74
BIC(lmer.slope.final.SelfDeclare.int)
# 10179.34

# Significant interaction effect between selfdelcare and X16minus17.  AIC stays about the same BIC gets worse but we ke

########### Problem 4
```

```
#a
lmer.popular <- lmer(Popular ~ as.factor(Harmony) + as.factor(Instrument) +
                          as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice))
summary(lmer.popular)
AIC(lmer.popular)
# 10097.24
BIC(lmer.popular)
# 10167.09


lmer.popular.2 <- lmer(Popular ~ as.factor(Harmony) + as.factor(Instrument) + (1|Subject:Instrument) + (1|Subject:Harmo
lmer.popular.3 <- lmer(Popular ~ as.factor(Harmony) + as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) +
lmer.popular.4 <- lmer(Popular ~ as.factor(Instrument) + as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony
lmer.popular.5 <- lmer(Popular ~ as.factor(Harmony) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice),
lmer.popular.6 <- lmer(Popular ~ as.factor(Voice) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), d
lmer.popular.7 <- lmer(Popular ~ as.factor(Instrument) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voic

list.models.lmer.all <- list(lmer.popular, lmer.popular.2,lmer.popular.3,lmer.popular.4,lmer.popular.5,lmer.popular.6,lr
lapply(list.models.lmer.all, AIC)
lapply(list.models.lmer.all, BIC)


#b
lmer.popular.pop.1 <- lmer(Popular ~  KnowRobOrAxis +X16.minus.17 + log.omsi + Selfdeclare2  + as.factor(Harmony) + as.
+ (1|Subject:Voice))
summary(lmer.popular.pop.1)
AIC(lmer.popular.pop.1)
# 10099.63
BIC(lmer.popular.pop.1)
# 10192.77

#AIC and BIC is slightly worse with more covariates (although some are significant).  We chose to elave them in but we

#c
lmer.popular.pop.1 <- lmer(Popular ~  Selfdeclare2*KnowRobOrAxis +
X16.minus.17 + log.omsi + Selfdeclare2  + as.factor(Harmony) + as.factor(Instrument) + as.factor(Voice) + (1|Subject:In
+ (1|Subject:Voice))
summary(lmer.popular.pop.1)
AIC(lmer.popular.pop.1)
# 10097.18
BIC(lmer.popular.pop.1)
# 10196.14


summary(lmer.popular.pop.1)
AIC(lmer.popular.pop.1)
# 10097.18
BIC(lmer.popular.pop.1)
# 10196.14

detach(ratings2)
```