1 a 9/9
b 7/9
c 6/9

2 a 5/9
b 9/9
c 9/9

3 7/9

4 a 9/9
b 9/9
c 9/9

5 10/10

Total 89/100

Hierarchical Models HW #5

Lauren Miller

December 18, 2015

you appear to have a pretty good grasp of things, but you have included far too much raw output, which you cannot expect a reader to wade through. It is difficult, but necessarty, to decide what is truly important to make your points, and then relegate the rest, if you must, to appendices.

1.

First we will load in the data and remove some rows with invalid values for the response variables. Then we will fit a linear model to predict classical music ratings.

a.

```
# Load in the data
setwd("~/MSP 15-16/Hierarchical Models")
rate <- read.csv("ratings.csv")

# EDA
table(rate$Classical)
```

```
##
##    0   1   2   3 3.5   4 4.2 4.6   5   6   7   8   9 9.5  10  19
##    8 111 231 254   1 239   1   1 297 266 334 305 203   1 240   1
```

```
table(rate$Popular)
```

```
##
##    0   1   2   3 3.5   4 4.2 4.6   5   6 6.8   7   8   9  10  19
##   25 171 197 218   1 282   1   1 371 314   1 350 312 136 112   1
```

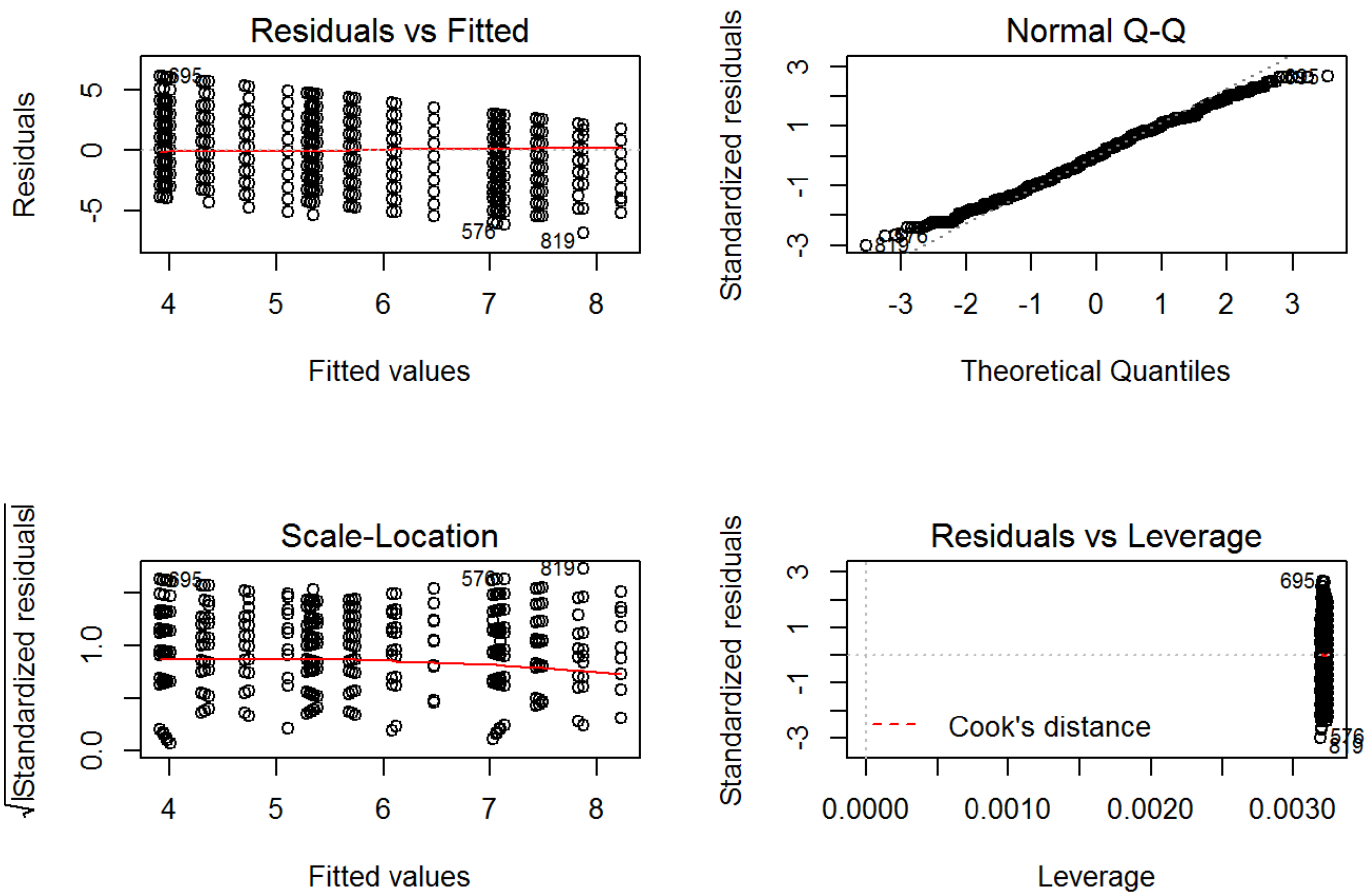
```
# Remove rows with 19 as a rating
rate <- rate[-which(rate$Classical == 19),]
rate <- rate[-which(rate$Popular == 19),]

# Fit linear regression to predict classical ratings from instrument, harmony, and voice
mod1 <- lm(Classical ~ Instrument + Harmony + Voice, data = rate)
summary(mod1)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8671 -1.7093 -0.0258  1.7762  6.0936
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.33534    0.12928  33.535  < 2e-16 ***
## Instrumentpiano   1.37395    0.11249  12.214  < 2e-16 ***
## Instrumentstring  3.11941    0.11181  27.899  < 2e-16 ***
## HarmonyI-V-IV    -0.03065    0.12952  -0.237 0.812942
## HarmonyI-V-VI     0.76909    0.12947   5.940 3.24e-09 ***
## HarmonyIV-I-V     0.03163    0.12942   0.244 0.806911
## Voicepar3rd      -0.39833    0.11225  -3.549 0.000394 ***
## Voicepar5th      -0.35673    0.11215  -3.181 0.001486 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.286 on 2483 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.2551, Adjusted R-squared:  0.253
## F-statistic: 121.5 on 7 and 2483 DF,  p-value: < 2.2e-16
```

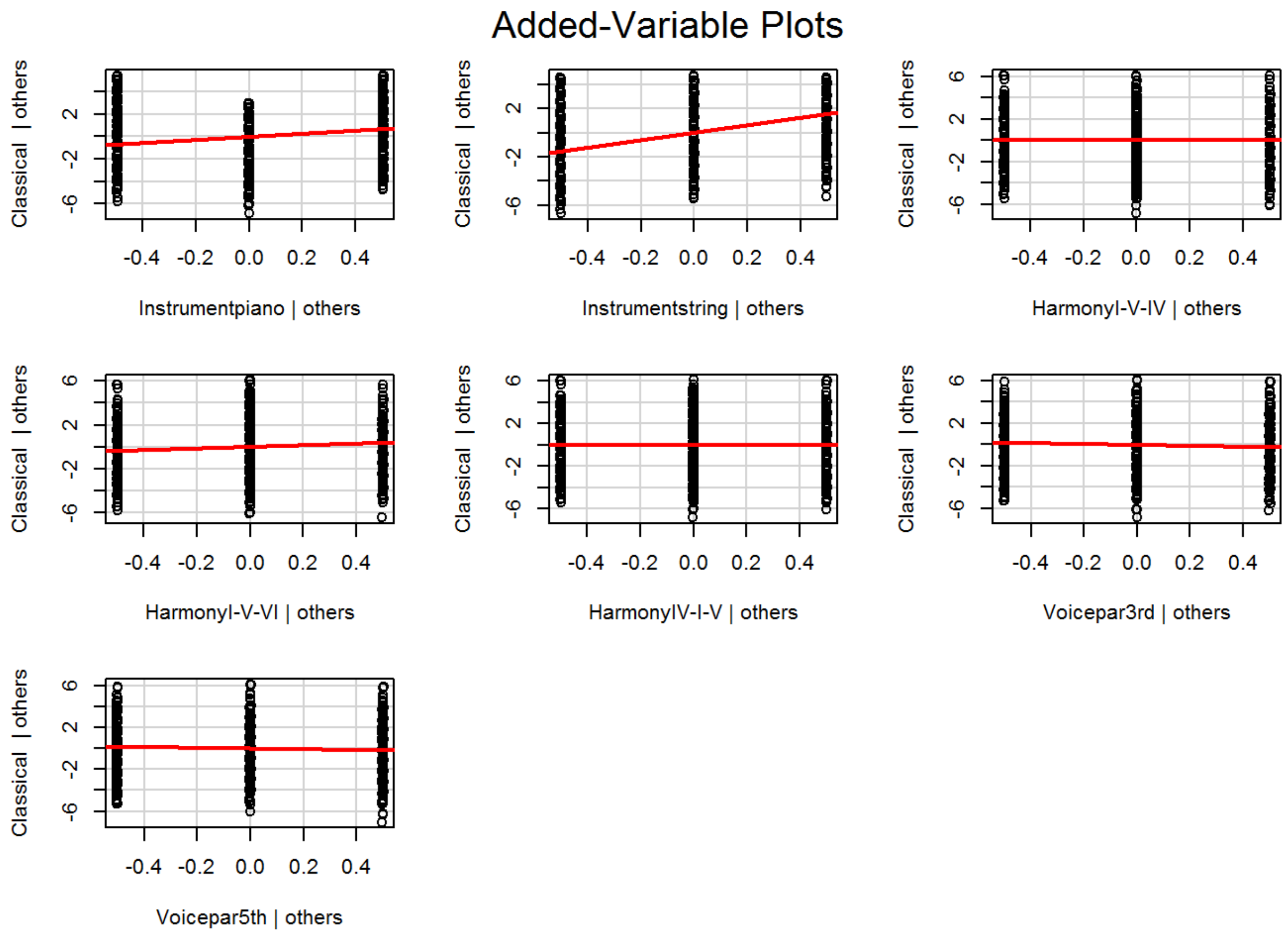
```
AIC(mod1)
```

```
## [1] 11198.2
```

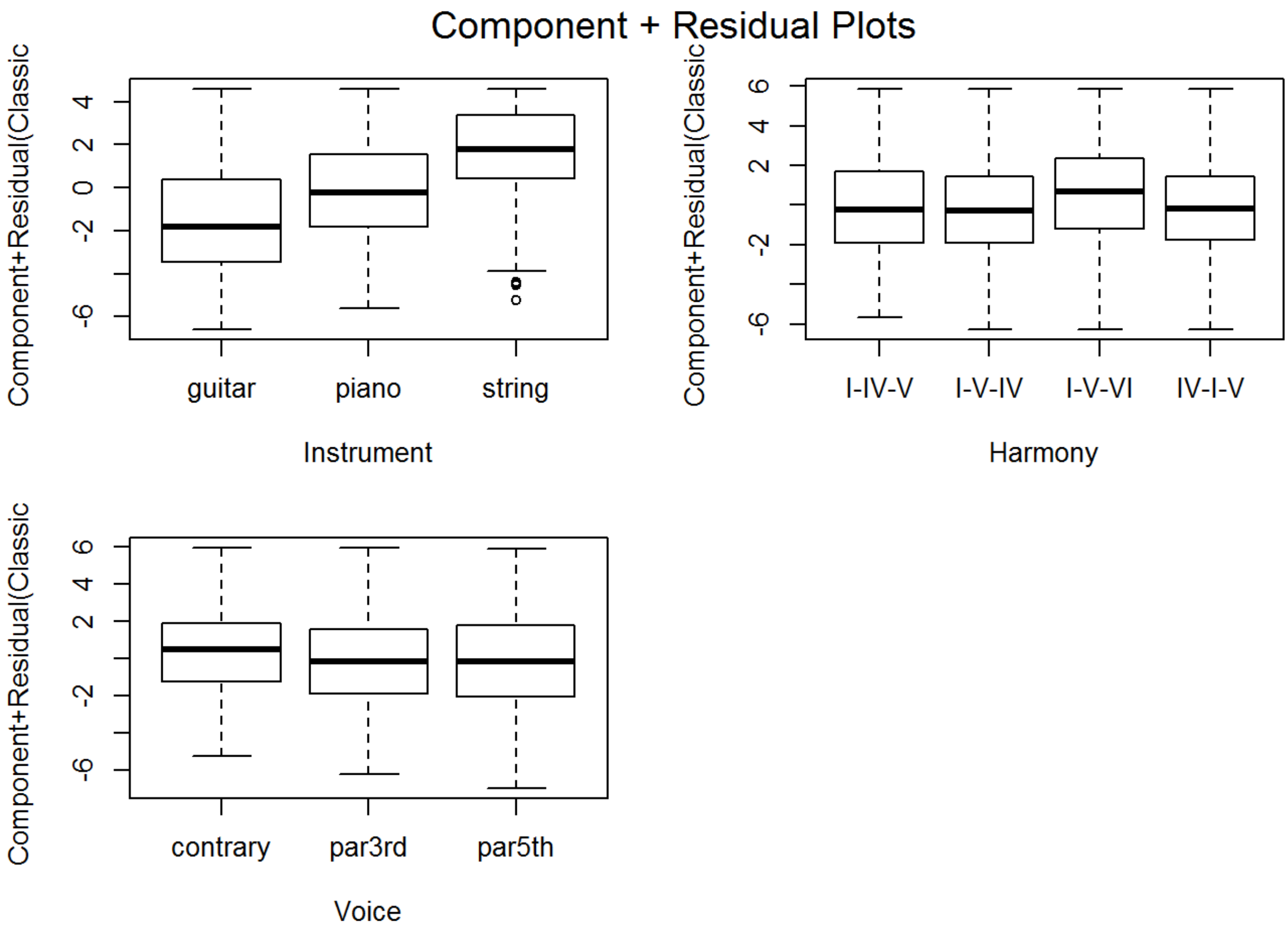
```
# Check model diagnostics
suppressMessages(library(car))
suppressMessages(library(arm))
par(mfrow=c(2,2))
plot(mod1)
```



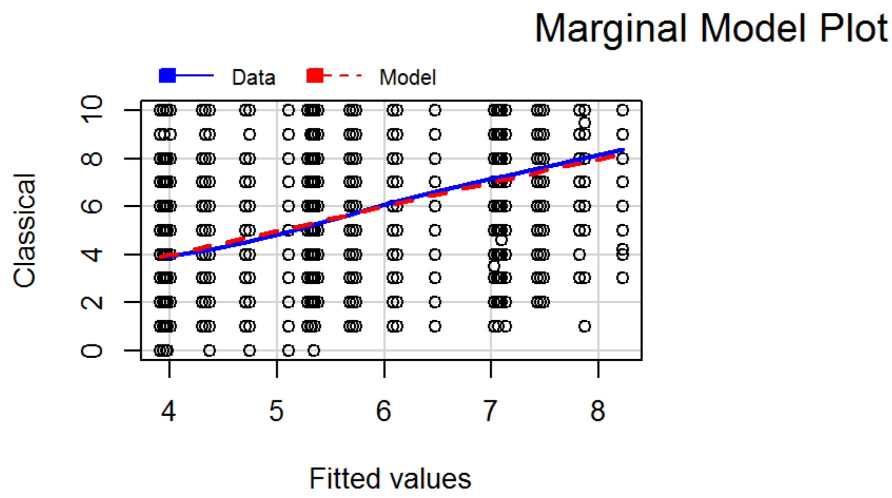

```
avPlots(mod1)
```




```
crPlots(mod1)
```




```
marginalModelPlots(mod1)
```



From the output of our linear model to predict classical ratings from the three experimental factors above, we can see that almost all of the categorical variables are significant in predicting classical ratings. The adjusted R squred of the model is 0.253 and the AIC is 11198.2. We will compare these values to those of models which each of the three variables to deterine if they are actually influential and need to be included in the model. We also check the model diagnostics to make sure that the linear model assumptions are met so that we can report on the p values and significance of the coefficients. The residuals have a little bit of a downward pattern toward higher fitted vaues. The QQ plot looks good, and all the points have leverage under 0.004. The added variable plots, component + residual plots, and marginal model plots all look like we would hope for a successful model.

```
# Fit linear regression to predict classical ratings from harmony and voice
mod2 <- lm(Classical ~ Harmony + Voice, data = rate)
summary(mod2)
```

```
##
## Call:
## lm(formula = Classical ~ Harmony + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6063 -2.2514  0.1656  2.1656  4.5876
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.83441    0.12838  45.445  < 2e-16 ***
## HarmonyI-V-IV -0.02763    0.14847  -0.186   0.8524
## HarmonyI-V-VI  0.77191    0.14841   5.201 2.14e-07 ***
## HarmonyIV-I-V  0.03145    0.14835   0.212   0.8321
## Voicepar3rd   -0.39442    0.12867  -3.065   0.0022 **
## Voicepar5th   -0.35495    0.12855  -2.761   0.0058 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.62 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.02047,    Adjusted R-squared:  0.0185
## F-statistic: 10.38 on 5 and 2485 DF,  p-value: 7.102e-10
```

```
AIC(mod2)
```

```
## [1] 11876.37
```

```
# Fit linear regression to predict classical ratings from instrument and voice
mod3 <- lm(Classical ~ Instrument + Voice, data = rate)
summary(mod3)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2919 -1.5457 -0.1283  1.7519  5.8717
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.5274     0.1033  43.819  < 2e-16 ***
## Instrumentpiano    1.3736     0.1136  12.090  < 2e-16 ***
## Instrumentstring   3.1198     0.1129  27.626  < 2e-16 ***
## Voicepar3rd       -0.3992     0.1134  -3.521 0.000438 ***
## Voicepar5th       -0.3553     0.1133  -3.137 0.001726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.309 on 2486 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.2392, Adjusted R-squared:  0.238
## F-statistic: 195.4 on 4 and 2486 DF,  p-value: < 2.2e-16
```

```
AIC(mod3)
```

```
## [1] 11244.82
```

```
# Fit linear regression to predict classical ratings from instrument and harmony
mod4 <- lm(Classical ~ Instrument + Harmony, data = rate)
summary(mod4)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9718 -1.8528 -0.0842  1.7851  5.9474
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.08419    0.11226  36.380  < 2e-16 ***
## Instrumentpiano    1.37382    0.11279  12.181  < 2e-16 ***
## Instrumentstring   3.11894    0.11211  27.821  < 2e-16 ***
## HarmonyI-V-IV     -0.03163    0.12987  -0.244    0.808
## HarmonyI-V-VI      0.76865    0.12981   5.921 3.64e-09 ***
## HarmonyIV-I-V      0.03160    0.12976   0.244    0.808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.292 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.2506, Adjusted R-squared:  0.249
## F-statistic: 166.2 on 5 and 2485 DF,  p-value: < 2.2e-16
```

```
AIC(mod4)
```

```
## [1] 11209.43
```

For our model without instrument, the adjusted R squared value is 0.0185 and the AIC is 11876.37. For our model without harmony, the adjusted R squared is 0.238 and the AIC is 11244.82. For the linear model without voice leading, the adjusted R squared is 0.249 and the AIC is 11209.43. We can see that
9 our model with all three predictors is the best fitting model in terms of AIC and adjusted R squared, so we will include all the predictors in our future models.

b. i.

$$y_i = \alpha_{j[i]} + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2)$$

$$\alpha_j = \beta_0 + \eta_j,\ \eta_j \sim N(0, \tau^2)$$

how should this be adjusted to accommodate the three experimental design factors?

where y is the classical ratings and j are the different participants.

ii.

```
# Fit random intercept model to predict classical ratings
mod5 <- lmer(Classical ~ 1 + (1|Subject), data = rate, REML = FALSE)
summary(mod5)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + (1 | Subject)
##    Data: rate
## 
##      AIC      BIC   logLik deviance df.resid 
##  11430.2  11447.7  -5712.1  11424.2     2488 
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.88420 -0.75277 -0.02021  0.75847  2.80133 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.620    1.273   
##  Residual             5.361    2.315   
## Number of obs: 2491, groups:  Subject, 70
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   5.7818     0.1591   36.34
```

```
# Test if random intercept is necessary with AIC/BIC/DIC
anova(mod5, mod1)
```

```
## Data: rate
## Models:
## mod5: Classical ~ 1 + (1 | Subject)
## mod1: Classical ~ Instrument + Harmony + Voice
##      Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)    
## mod5  3 11430 11448 -5712.1    11424                             
## mod1  9 11198 11251 -5590.1    11180 244.03      6  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above table, we can see that AIC and BIC prefer the linear model without the random intercept. We also want to test our models using simulation of the likelihood ratio test.

```
# Simulate LRT test
library(RLRsim)
```

```
## Warning: package 'RLRsim' was built under R version 3.2.3
```

```
exactRLRT(mod5)
```

```
## Using restricted likelihood evaluated at ML estimators.
## Refit with method="REML" for exact results.
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 491.85, p-value < 2.2e-16
```

From the above output of the simulation of the REML likelihood ratio test, we can see that the p value is 2.2x10^-16, so we reject the null hypothesis and keep the random effect. These two tests are telling different stories, but it is probably because the linear model also includes three other predictive variables. We will now include those as well and test if the random effect is needed.

```
# Add the 3 variables into the random intercept model
mod6 <- lmer(Classical ~ Instrument + Harmony + Voice + (1|Subject), data = rate, REML =
FALSE)
exactRLRT(mod6)
```

```
## Using restricted likelihood evaluated at ML estimators.
## Refit with method="REML" for exact results.
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 769.45, p-value < 2.2e-16
```

Again, we can see that there is a small p value, so we still find the random effect to be useful in predicting the classical ratings.

iii.

```
# Output summary of random intercept model with 3 fixed effects
summary(mod6)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: rate
## 
##      AIC      BIC   logLik deviance df.resid 
##  10430.5  10488.7  -5205.3  10410.5     2481 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.0104 -0.6429 -0.0136  0.6485  3.9132 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.669    1.292   
##  Residual             3.527    1.878   
## Number of obs: 2491, groups:  Subject, 70
## 
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        4.33811    0.18743   23.15
## Instrumentpiano    1.37553    0.09251   14.87
## Instrumentstring   3.11874    0.09190   33.94
## HarmonyI-V-IV     -0.03005    0.10642   -0.28
## HarmonyI-V-VI      0.77098    0.10637    7.25
## HarmonyIV-I-V      0.03268    0.10633    0.31
## Voicepar3rd       -0.40032    0.09222   -4.34
## Voicepar5th       -0.36102    0.09214   -3.92
## 
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.244                                                 
## Instrmntstr -0.246  0.498                                          
## HrmnyI-V-IV -0.282  0.002    -0.001                                
## HrmnyI-V-VI -0.283  0.001    -0.001     0.499                      
## HrmnyIV-I-V -0.283 -0.001     0.000     0.499  0.499               
## Voicepar3rd -0.246  0.000    -0.001    -0.001  0.001  0.001        
## Voicepar5th -0.245 -0.001    -0.001    -0.002 -0.003 -0.002  0.500
```

From the above model output where we added fixed effects for instrument, harmony, and voice leading,
we can see that AIC and BIC prefer this model. That implies that at least one of instrument, harmony, or
7 voice leading are influential in predicting the classical rating. There are large t values on some of the
coefficients, such as for instruments and voice leading, so these variables are significant and should be
included in our model.

c. i.

```
# Fit model with three random effect terms for each key variable
mod7 <- lmer(Classical ~ 1 + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Vo
ice), data = rate, REML = FALSE)
summary(mod7)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ 1 + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##  10173.6  10202.7  -5081.8  10163.6     2486
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4559 -0.5547  0.0230  0.5322  4.3440
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.55370  0.7441
##  Subject:Voice      (Intercept) 0.05723  0.2392
##  Subject:Instrument (Intercept) 3.90182  1.9753
##  Residual                       2.40669  1.5514
## Number of obs: 2491, groups:
## Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   5.7752     0.1477   39.11
```

```
# Compare linear regression, random intercept model with fixed effects, and model with 3
ranodm components
anova(mod7, mod6, mod1)
```

```
## Data: rate
## Models:
## mod7: Classical ~ 1 + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
## mod7:     (1 | Subject:Voice)
## mod1: Classical ~ Instrument + Harmony + Voice
## mod6: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##      Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod7  5 10174 10203 -5081.8    10164
## mod1  9 11198 11251 -5590.1    11180   0.00      4          1
## mod6 10 10430 10489 -5205.3    10410 769.69      1     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the summary above where we built a model with three random effects for each of the experiemental design factors,we can see that the AIC is 10173.6 and the BIC is 10202.7, so this model is preferred to the previous model where we treated the covariates as random effects. The AIC is also lower than the AIC of the linear regression, so we also prefer this new model.

ii.

```
# Examine the influence of the random effect on voice
exactRLRT(lmer(Classical ~ 1 + (1|Subject:Voice), data = rate), m0 = lmer(Classical ~ 1 +
(1|Subject:Instrument), data = rate), mA = lmer(Classical ~ 1 + (1|Subject:Instrument) +
(1|Subject:Voice), data = rate))
```

why are the experiemental design factors not included as fixed effects in these models?

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 0.49844, p-value = 0.2266
```

```
# Examine the influence of the random effect on harmony
exactRLRT(lmer(Classical ~ 1 + (1|Subject:Harmony), data = rate), m0 = lmer(Classical ~ 1
+ (1|Subject:Instrument), data = rate), mA = lmer(Classical ~ 1 + (1|Subject:Instrument)
+ (1|Subject:Harmony), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 150.58, p-value < 2.2e-16
```

```
# Examine the influence of the random effect on instrument
exactRLRT(lmer(Classical ~ 1 + (1|Subject:Instrument), data = rate), m0 = lmer(Classical
~ 1 + (1|Subject:Harmony), data = rate), mA = lmer(Classical ~ 1 + (1|Subject:Instrument)
+ (1|Subject:Harmony), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 1393.4, p-value < 2.2e-16
```

We ran simulations of likelihod ratio tests to examine the influence of each of the random effects. The p value for the LRT for voice is 0.2346, so this is not significant ad not improving the model fit. The p value from the LRT with harmony is less than 0.01, so it is significant and we will include it. The p value for the
6 LRT on instrument is also less than 0.01, so we wil include it as well.

ok, but you have not tested the random effects in the model intended.

We can see from the summary ouput of this model that the variance component on voice leading is very small at 0.057. For harmony is is 0.554 and for instrument it is the largest at 3.902. The residual variance is 2.407, which is smaller that in the previous model with the fixed effects. This means that this model is fitting the best so far.

iii.

$$y_i = \alpha^{instrument}_{j[i]k[i]} + \alpha^{harmony}_{j[i]l[i]} + \alpha^{voice}_{j[i]m[i]} + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2)$$

$$\alpha^{instrument}_{jk} = \beta_{0k} + \eta_{jk},\ \eta_{jk} \sim N(0, \tau^2_{instrument})$$

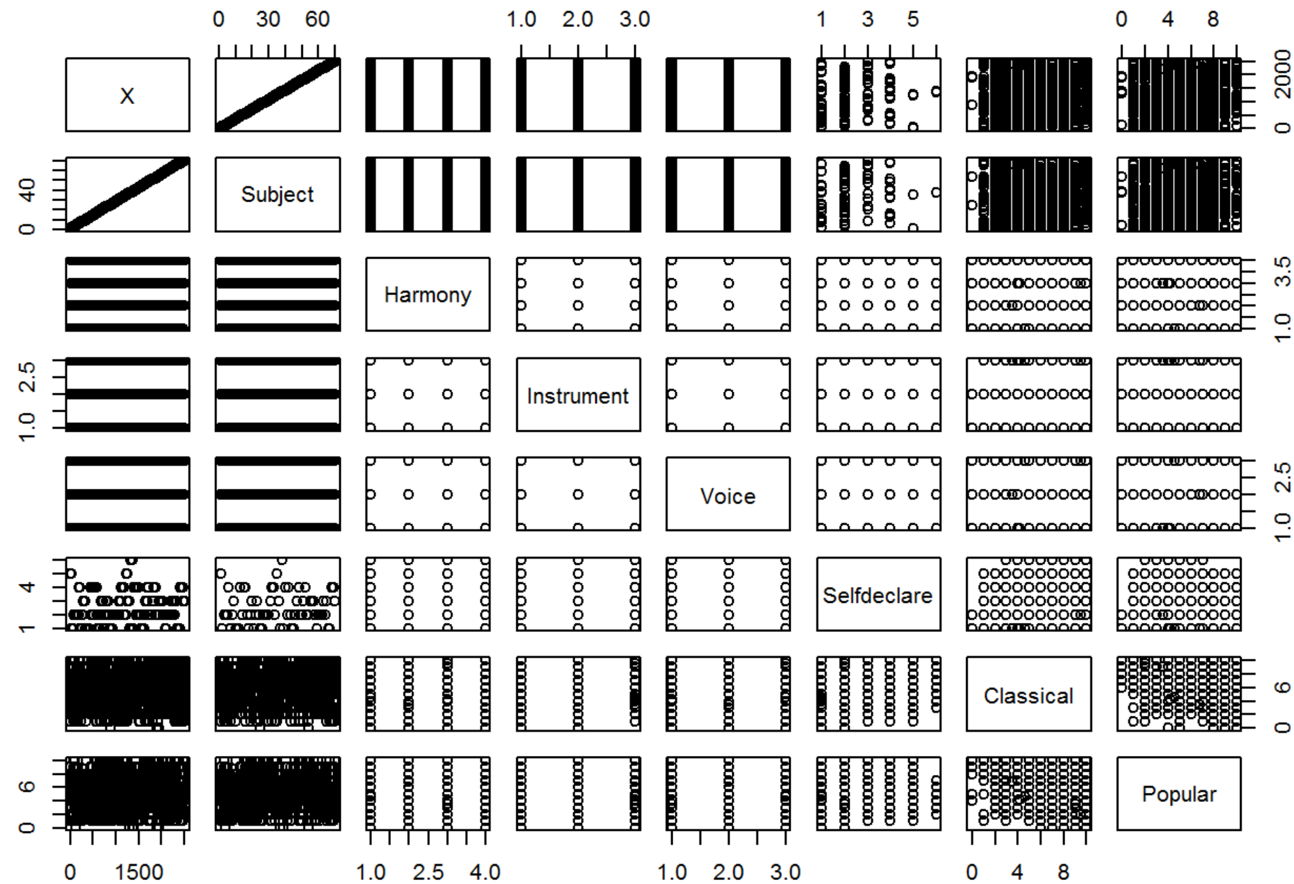
$$\alpha^{harmony}_{jl} = \beta_{0l} + \eta_{jl},\ \eta_{jl} \sim N(0, \tau^2_{harmony})$$

$$\alpha^{voice}_{jm} = \beta_{0m} + \eta_{jm},\ \eta_{jm} \sim N(0, \tau^2_{voice})$$

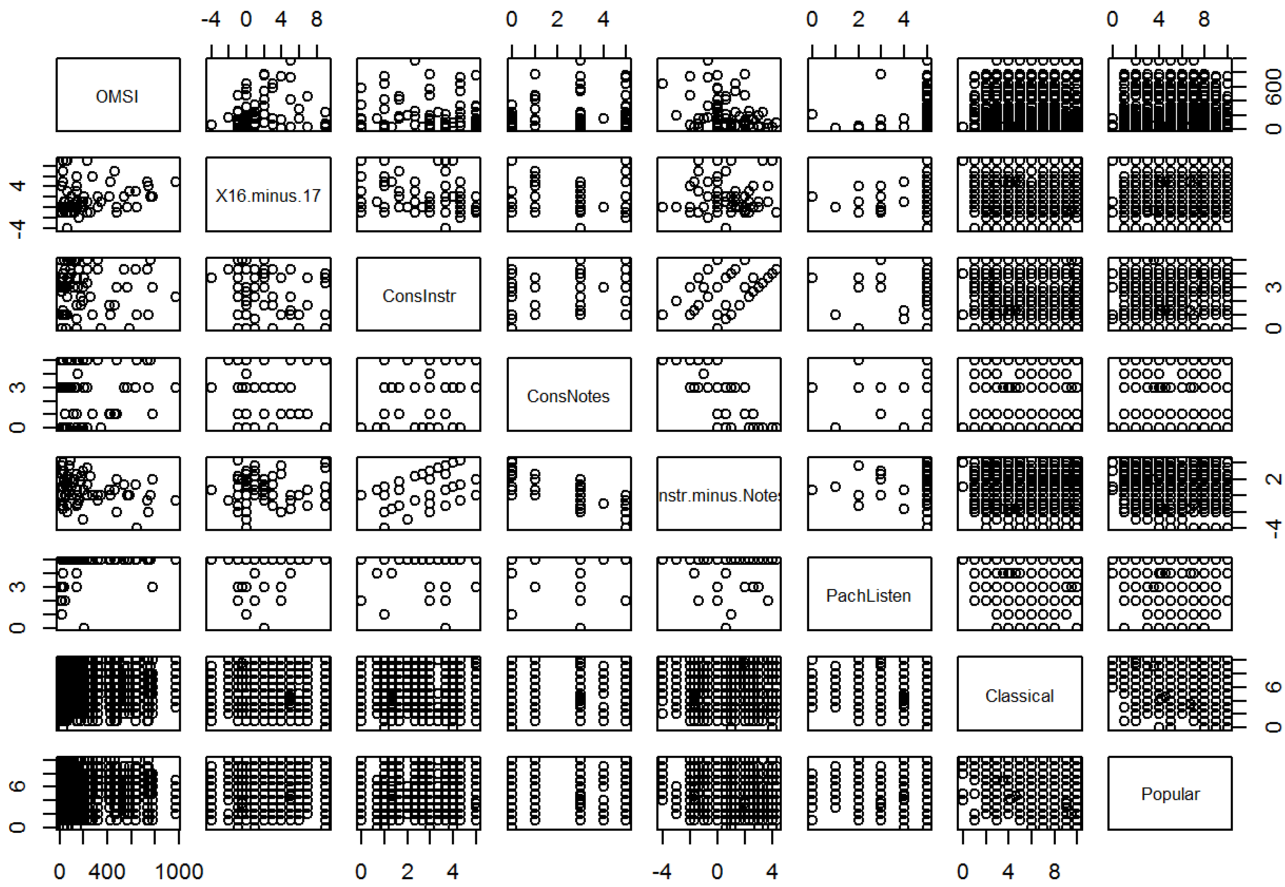
2.

a.

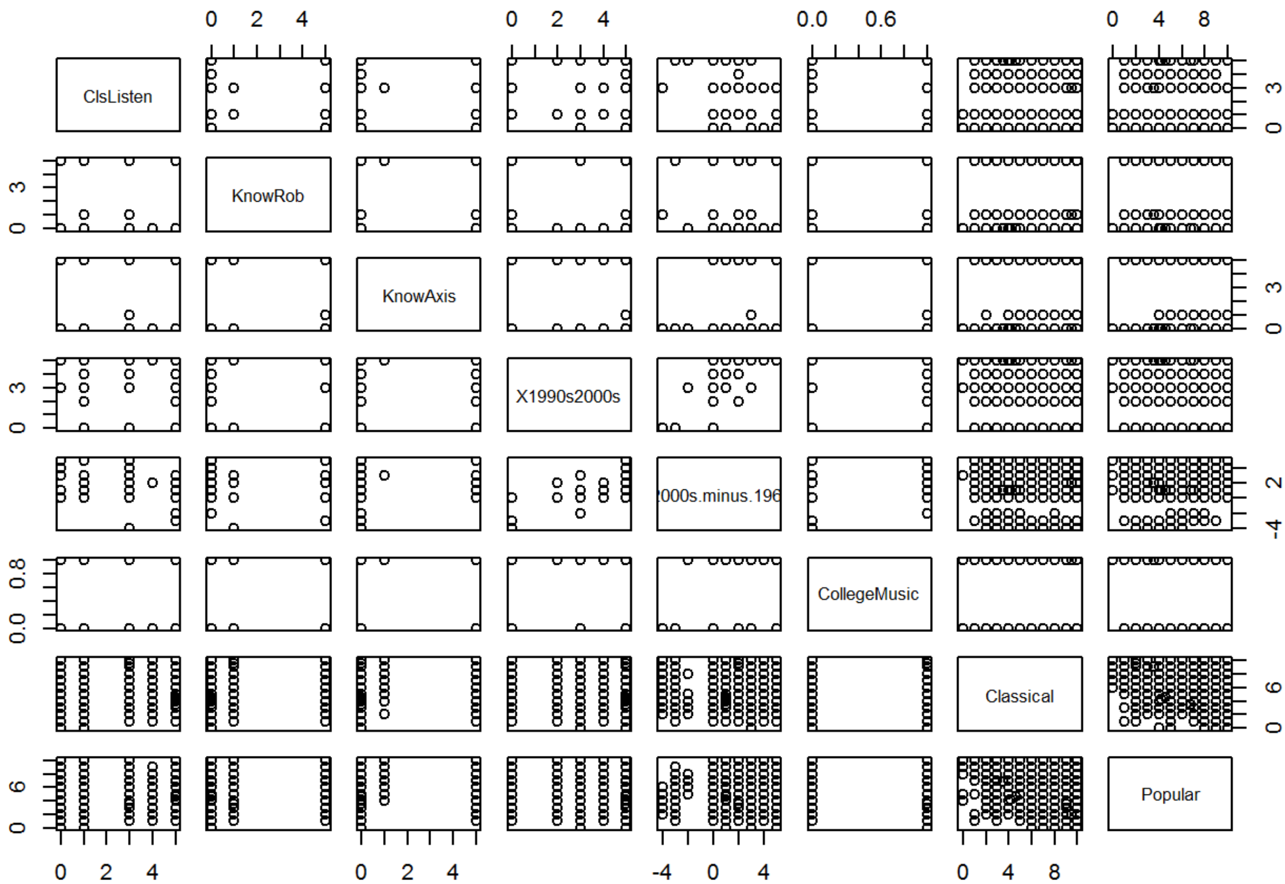
```
# Look at the relatiionships between variables
pairs(rate[,c(1:6, 27, 28)])
```



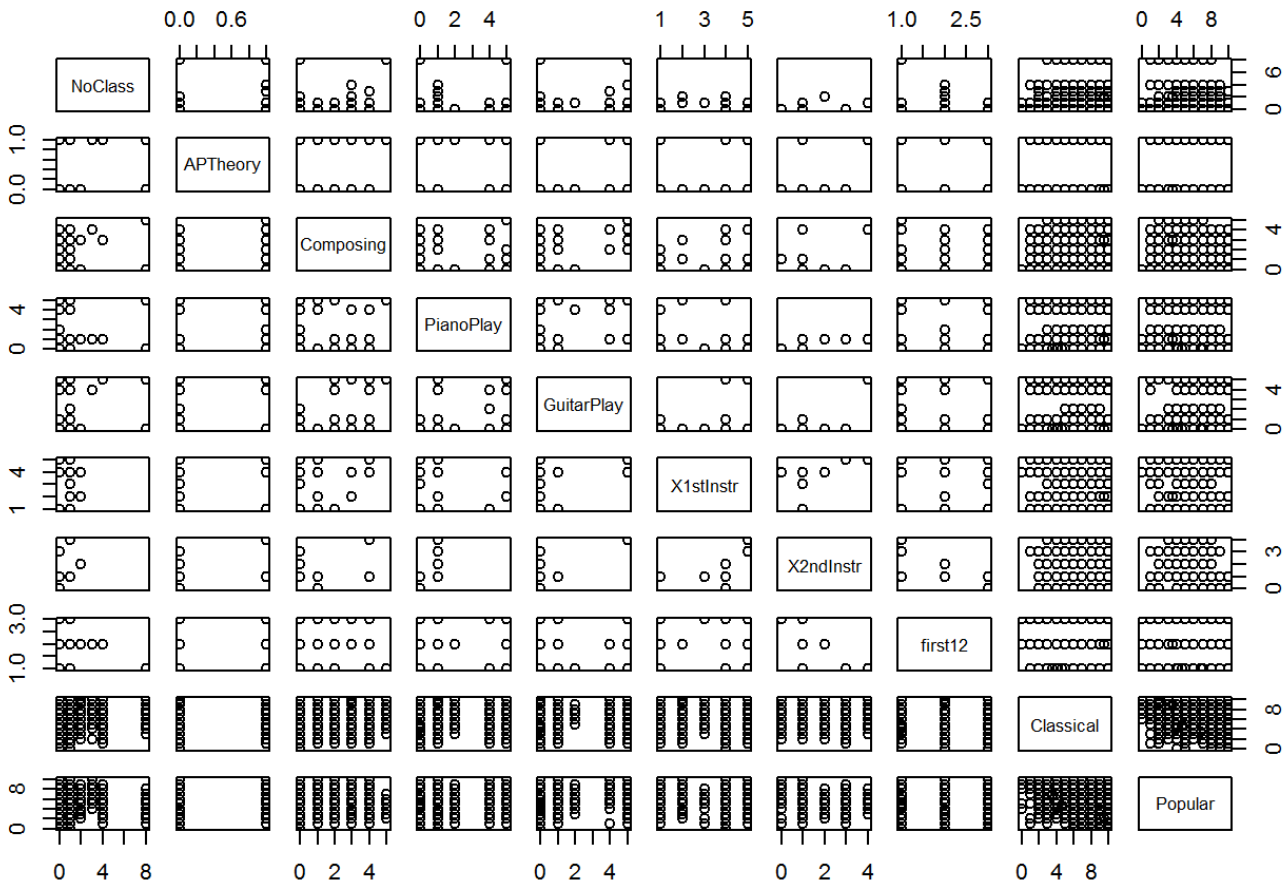

```
pairs(rate[,c(7:12, 27, 28)])
```




```
pairs(rate[,c(13:18, 27, 28)])
```




```
pairs(rate[,19:28])
```




```
# Look at distributions of chosen variables
par(mfrow = c(3, 2))
hist(rate$OMSI)
hist(rate$PachListen)
hist(rate$X1990s2000s)
hist(rate$NoClass)
hist(rate$Composing)

# Fit multileve model with fixed effects for other covariates
# Add OMSI
mod8 <- lmer(Classical ~ 1 + OMSI + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Sub
ject:Voice), data = rate, REML = FALSE)
summary(mod8)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ 1 + OMSI + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##  10175.6  10210.5  -5081.8  10163.6     2485
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4559 -0.5544  0.0229  0.5316  4.3444
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.55370  0.7441
##  Subject:Voice      (Intercept) 0.05723  0.2392
##  Subject:Instrument (Intercept) 3.90154  1.9752
##  Residual                       2.40669  1.5514
## Number of obs: 2491, groups:
## Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 5.7582390  0.2064351  27.894
## OMSI        0.0000751  0.0006386   0.118
##
## Correlation of Fixed Effects:
##      (Intr)
## OMSI -0.699
```

```
# Add PachListen
mod9 <- lmer(Classical ~ 1 + factor(PachListen) + (1|Subject:Instrument) + (1|Subject:Har
mony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod9)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + factor(PachListen) + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   9904.9   9962.9  -4942.5   9884.9     2421
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4858 -0.5497  0.0195  0.5294  4.3927
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.57783  0.7602
##  Subject:Voice      (Intercept) 0.05458  0.2336
##  Subject:Instrument (Intercept) 3.80580  1.9508
##  Residual                       2.36976  1.5394
## Number of obs: 2431, groups:
## Subject:Harmony, 272; Subject:Voice, 204; Subject:Instrument, 204
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)           5.6667     1.2236   4.631
## factor(PachListen)1   0.8056     1.7304   0.466
## factor(PachListen)2   0.1319     1.3680   0.096
## factor(PachListen)3  -1.0028     1.3403  -0.748
## factor(PachListen)4  -0.3014     1.4985  -0.201
## factor(PachListen)5   0.1787     1.2346   0.145
##
## Correlation of Fixed Effects:
##             (Intr) f(PL)1 f(PL)2 f(PL)3 f(PL)4
## fctr(PchL)1 -0.707
## fctr(PchL)2 -0.894  0.632
## fctr(PchL)3 -0.913  0.645  0.816
## fctr(PchL)4 -0.816  0.577  0.730  0.745
## fctr(PchL)5 -0.991  0.701  0.886  0.905  0.809
```

```
# Add x1990s2000s
mod10 <- lmer(Classical ~ 1 + factor(X1990s2000s) + (1|Subject:Instrument) + (1|Subject:H
armony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod10)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + factor(X1990s2000s) + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   9640.3   9692.2  -4811.1   9622.3     2350
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4655 -0.5528  0.0232  0.5333  4.3350
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.54013  0.7349
##  Subject:Voice      (Intercept) 0.06289  0.2508
##  Subject:Instrument (Intercept) 3.89897  1.9746
##  Residual                       2.40630  1.5512
## Number of obs: 2359, groups:
## Subject:Harmony, 264; Subject:Voice, 198; Subject:Instrument, 198
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)            6.2083     0.5037  12.325
## factor(X1990s2000s)2  -0.9306     0.8725  -1.067
## factor(X1990s2000s)3  -0.3904     0.6503  -0.600
## factor(X1990s2000s)4  -0.7972     0.7472  -1.067
## factor(X1990s2000s)5  -0.3920     0.5378  -0.729
##
## Correlation of Fixed Effects:
##                (Intr) f(X19902000)2 f(X19902000)3 f(X19902000)4
## f(X19902000)2 -0.577
## f(X19902000)3 -0.775  0.447
## f(X19902000)4 -0.674  0.389         0.522
## f(X19902000)5 -0.937  0.541         0.726         0.632
```

```
# Add Composing
mod11 <- lmer(Classical ~ 1 + factor(Composing) + (1|Subject:Instrument) + (1|Subject:Har
mony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod11)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + factor(Composing) + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   9852.6   9910.5  -4916.3   9832.6     2409
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5122 -0.5462  0.0249  0.5300  3.4573
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.57100  0.7556
##  Subject:Voice      (Intercept) 0.05322  0.2307
##  Subject:Instrument (Intercept) 3.83715  1.9589
##  Residual                       2.36437  1.5377
## Number of obs: 2419, groups:
## Subject:Harmony, 272; Subject:Voice, 204; Subject:Instrument, 204
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)          5.6186     0.1916  29.324
## factor(Composing)1   0.3906     0.5026   0.777
## factor(Composing)2   0.8455     0.4753   1.779
## factor(Composing)3  -0.5595     0.6427  -0.871
## factor(Composing)4   0.8026     0.5018   1.600
## factor(Composing)5  -0.2575     1.2417  -0.207
##
## Correlation of Fixed Effects:
##             (Intr) fc(C)1 fc(C)2 fc(C)3 fc(C)4
## fctr(Cmps)1 -0.381
## fctr(Cmps)2 -0.403  0.154
## fctr(Cmps)3 -0.298  0.114  0.120
## fctr(Cmps)4 -0.382  0.146  0.154  0.114
## fctr(Cmps)5 -0.154  0.059  0.062  0.046  0.059
```

```
# Add NoClass
mod12 <- lmer(Classical ~ 1 + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmo
ny) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod12)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + factor(NoClass) + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   9042.8   9099.8  -4511.4   9022.8     2205
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4813 -0.5479  0.0224  0.5219  3.4591
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.60246  0.7762
##  Subject:Voice      (Intercept) 0.05795  0.2407
##  Subject:Instrument (Intercept) 3.97641  1.9941
##  Residual                       2.37120  1.5399
## Number of obs: 2215, groups:
## Subject:Harmony, 248; Subject:Voice, 186; Subject:Instrument, 186
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)        5.6167     0.2452  22.908
## factor(NoClass)1   0.3058     0.3349   0.913
## factor(NoClass)2  -0.2070     0.9169  -0.226
## factor(NoClass)3   1.4993     1.2742   1.177
## factor(NoClass)4  -0.9778     1.2733  -0.768
## factor(NoClass)8  -0.1028     0.9169  -0.112
##
## Correlation of Fixed Effects:
##             (Intr) f(NC)1 f(NC)2 f(NC)3 f(NC)4
## fctr(NCls)1 -0.732
## fctr(NCls)2 -0.267  0.196
## fctr(NCls)3 -0.192  0.141  0.051
## fctr(NCls)4 -0.193  0.141  0.051  0.037
## fctr(NCls)8 -0.267  0.196  0.072  0.051  0.051
```

```
# Add multiple covariates as fixed effects at once
# Add PachListen and x1990s2000s
mod13 <- lmer(Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + (1|Subject:Instr
ument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod13)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + (1 |
##     Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
## 
##      AIC      BIC   logLik deviance df.resid 
##   9484.4   9564.9  -4728.2   9456.4     2309 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -4.4789 -0.5513  0.0211  0.5280  4.3669 
## 
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.55393  0.7443  
##  Subject:Voice      (Intercept) 0.06267  0.2503  
##  Subject:Instrument (Intercept) 3.79264  1.9475  
##  Residual                       2.38508  1.5444  
## Number of obs: 2323, groups:  
## Subject:Harmony, 260; Subject:Voice, 195; Subject:Instrument, 195
## 
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)            6.06411    1.33505   4.542
## factor(PachListen)1    0.80556    1.72617   0.467
## factor(PachListen)2    0.08831    1.38136   0.064
## factor(PachListen)3   -0.71777    1.38819  -0.517
## factor(PachListen)4   -0.32605    1.51244  -0.216
## factor(PachListen)5    0.15541    1.23696   0.126
## factor(X1990s2000s)2 -0.65068    0.89558  -0.727
## factor(X1990s2000s)3 -0.34812    0.65960  -0.528
## factor(X1990s2000s)4 -0.62035    0.75173  -0.825
## factor(X1990s2000s)5 -0.39744    0.54084  -0.735
## 
## Correlation of Fixed Effects:
##                (Intr) f(PL)1 f(PL)2 f(PL)3 f(PL)4 f(PL)5 f(X19902000)2
## fctr(PchL)1    -0.646                                                  
## fctr(PchL)2    -0.853  0.625                                           
## fctr(PchL)3    -0.814  0.622  0.793                                    
## fctr(PchL)4    -0.746  0.571  0.717  0.713                             
## fctr(PchL)5    -0.924  0.698  0.882  0.877  0.806                      
## f(X19902000)2 -0.219  0.000  0.050 -0.111  0.001  0.004              
## f(X19902000)3 -0.293  0.000  0.075  0.004 -0.089 -0.002  0.437        
## f(X19902000)4 -0.249  0.000 -0.015 -0.081  0.002  0.006  0.408        
## f(X19902000)5 -0.405  0.000  0.112  0.024  0.021  0.054  0.541        
##                f(X19902000)3 f(X19902000)4
## fctr(PchL)1                              
## fctr(PchL)2                              
## fctr(PchL)3                              
## fctr(PchL)4                              
```

```
## fctr(PchL)5
## f(X19902000)2
## f(X19902000)3
## f(X19902000)4  0.494
## f(X19902000)5  0.724         0.616
```

```
# Add Composing
mod14 <- lmer(Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + factor(Composin
g) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML
= FALSE)
summary(mod14)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + factor(Composing) +
##     (1 | Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   9164.6   9273.3  -4563.3   9126.6     2232
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5365 -0.5464  0.0212  0.5264  3.4595
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.57293  0.7569
##  Subject:Voice      (Intercept) 0.05878  0.2425
##  Subject:Instrument (Intercept) 3.75403  1.9375
##  Residual                       2.33871  1.5293
## Number of obs: 2251, groups:
## Subject:Harmony, 252; Subject:Voice, 189; Subject:Instrument, 189
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)           6.003992   1.342480   4.472
## factor(PachListen)1   0.805556   1.719961   0.468
## factor(PachListen)2   0.096887   1.378762   0.070
## factor(PachListen)3  -0.722105   1.400369  -0.516
## factor(PachListen)4   0.788132   1.812828   0.435
## factor(PachListen)5  -0.004686   1.238429  -0.004
## factor(X1990s2000s)2 -0.653394   0.936153  -0.698
## factor(X1990s2000s)3 -0.458791   0.784401  -0.585
## factor(X1990s2000s)4 -0.534419   0.751010  -0.712
## factor(X1990s2000s)5 -0.337325   0.568437  -0.593
## factor(Composing)1    0.513018   0.516372   0.994
## factor(Composing)2    0.805625   0.531757   1.515
## factor(Composing)3   -0.458656   0.680339  -0.674
## factor(Composing)4    0.844168   0.686475   1.230
## factor(Composing)5   -0.179404   1.338346  -0.134
##
## Correlation of Fixed Effects:
##                (Intr) f(PL)1 f(PL)2 f(PL)3 f(PL)4 f(PL)5 f(X19902000)2
## fctr(PchL)1    -0.641
## fctr(PchL)2    -0.852  0.624
## fctr(PchL)3    -0.790  0.614  0.779
## fctr(PchL)4    -0.615  0.474  0.595  0.588
## fctr(PchL)5    -0.906  0.694  0.874  0.873  0.675
## f(X19902000)2 -0.246  0.000  0.065 -0.116 -0.010 -0.006
## f(X19902000)3 -0.290  0.000  0.082 -0.006 -0.218 -0.008  0.444
## f(X19902000)4 -0.254  0.000 -0.013 -0.083  0.002  0.001  0.401
```

```
## f(X19902000)5 -0.423  0.000  0.125  0.008  0.017  0.039  0.581
## fctr(Cmps)1    0.029  0.000 -0.017 -0.102  0.021 -0.065 -0.116
## fctr(Cmps)2   -0.105  0.000  0.041 -0.026 -0.029 -0.065  0.199
## fctr(Cmps)3   -0.083  0.000  0.038 -0.127 -0.017 -0.043  0.200
## fctr(Cmps)4    0.020  0.000 -0.010 -0.029  0.103 -0.043 -0.003
## fctr(Cmps)5    0.005  0.000 -0.003 -0.012  0.121 -0.012 -0.008
##               f(X19902000)3 f(X19902000)4 f(X19902000)5 fc(C)1 fc(C)2
## fctr(PchL)1
## fctr(PchL)2
## fctr(PchL)3
## fctr(PchL)4
## fctr(PchL)5
## f(X19902000)2
## f(X19902000)3
## f(X19902000)4  0.430
## f(X19902000)5  0.684         0.599
## fctr(Cmps)1   -0.098         0.022        -0.068
## fctr(Cmps)2    0.246         0.071         0.247         0.117
## fctr(Cmps)3    0.181         0.016         0.195         0.108  0.186
## fctr(Cmps)4   -0.273         0.001        -0.047         0.146  0.079
## fctr(Cmps)5   -0.288         0.002        -0.012         0.089  0.021
##               fc(C)3 fc(C)4
## fctr(PchL)1
## fctr(PchL)2
## fctr(PchL)3
## fctr(PchL)4
## fctr(PchL)5
## f(X19902000)2
## f(X19902000)3
## f(X19902000)4
## f(X19902000)5
## fctr(Cmps)1
## fctr(Cmps)2
## fctr(Cmps)3
## fctr(Cmps)4    0.054
## fctr(Cmps)5    0.016  0.180
```

```
# Add NoClass
mod15 <- lmer(Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + factor(Composin
g) + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice),
data = rate, REML = FALSE)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
summary(mod15)
```

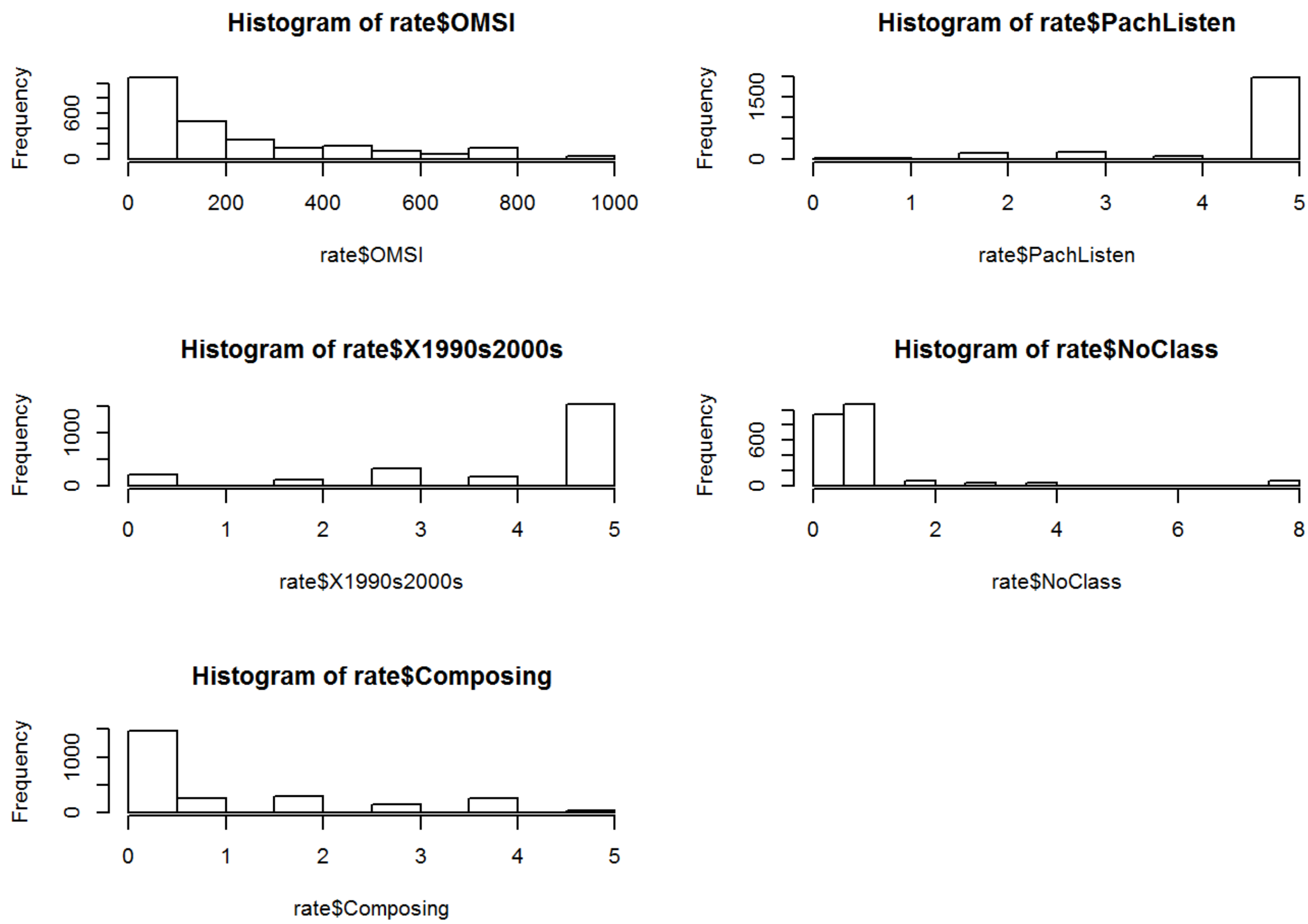
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Classical ~ 1 + factor(PachListen) + factor(X1990s2000s) + factor(Composing) +
##     factor(NoClass) + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8469.2   8598.9  -4211.6   8423.2     2048
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4666 -0.5488  0.0229  0.5244  3.4323
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.59048  0.7684
##  Subject:Voice      (Intercept) 0.06636  0.2576
##  Subject:Instrument (Intercept) 3.74549  1.9353
##  Residual                       2.36451  1.5377
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           4.87367    1.40361   3.472
## factor(PachListen)1   0.80556    1.72274   0.468
## factor(PachListen)2  -0.54728    1.43738  -0.381
## factor(PachListen)3  -1.12701    1.50449  -0.749
## factor(PachListen)4   0.80909    1.82386   0.444
## factor(PachListen)5   0.04413    1.26032   0.035
## factor(X1990s2000s)2  0.37348    1.13252   0.330
## factor(X1990s2000s)3  0.65057    0.88871   0.732
## factor(X1990s2000s)4  0.70118    0.98725   0.710
## factor(X1990s2000s)5  0.79299    0.69728   1.137
## factor(Composing)1    0.49463    0.53287   0.928
## factor(Composing)2    0.59431    0.60544   0.982
## factor(Composing)3   -0.03841    0.96320  -0.040
## factor(Composing)4    0.52467    0.81131   0.647
## factor(Composing)5   -0.20727    1.37176  -0.151
## factor(NoClass)1      0.12255    0.36223   0.338
## factor(NoClass)2      0.51346    1.13765   0.451
## factor(NoClass)3      0.88025    1.47861   0.595
## factor(NoClass)4     -1.03350    1.56911  -0.659
##
## Correlation of Fixed Effects:
##               (Intr) f(PL)1 f(PL)2 f(PL)3 f(PL)4 f(PL)5 f(X19902000)2
## fctr(PchL)1   -0.614
## fctr(PchL)2   -0.757  0.599
## fctr(PchL)3   -0.692  0.573  0.686
```

```
## fctr(PchL)4   -0.590  0.472  0.568  0.541
## fctr(PchL)5   -0.868  0.683  0.837  0.829  0.660
## f(X19902000)2 -0.331  0.000  0.040 -0.117 -0.004  0.014
## f(X19902000)3 -0.368  0.000  0.030 -0.018 -0.204  0.016  0.532
## f(X19902000)4 -0.268  0.000 -0.131  0.038 -0.001  0.025  0.306
## f(X19902000)5 -0.497  0.000  0.043 -0.022  0.022  0.058  0.666
## fctr(Cmps)1    0.048  0.000 -0.021 -0.157  0.023 -0.089 -0.156
## fctr(Cmps)2   -0.069  0.000  0.004 -0.029 -0.034 -0.043  0.137
## fctr(Cmps)3   -0.169  0.000  0.077 -0.174 -0.006 -0.022  0.428
## fctr(Cmps)4    0.020  0.000 -0.006 -0.026  0.133 -0.031 -0.021
## fctr(Cmps)5    0.012  0.000 -0.014 -0.042  0.130 -0.042 -0.019
## fctr(NCls)1    0.029  0.000 -0.064 -0.149  0.021 -0.169 -0.026
## fctr(NCls)2    0.093  0.000 -0.049 -0.128  0.003 -0.047 -0.405
## fctr(NCls)3    0.014  0.000 -0.012 -0.025 -0.085 -0.039 -0.001
## fctr(NCls)4    0.127  0.000 -0.062  0.070 -0.008 -0.039 -0.274
##               f(X19902000)3 f(X19902000)4 f(X19902000)5 fc(C)1 fc(C)2
## fctr(PchL)1
## fctr(PchL)2
## fctr(PchL)3
## fctr(PchL)4
## fctr(PchL)5
## f(X19902000)2
## f(X19902000)3
## f(X19902000)4  0.426
## f(X19902000)5  0.740         0.540
## fctr(Cmps)1   -0.123        -0.021        -0.096
## fctr(Cmps)2    0.179         0.134         0.139         0.107
## fctr(Cmps)3    0.278        -0.116         0.340         0.032  0.101
## fctr(Cmps)4   -0.305        -0.002        -0.040         0.138  0.056
## fctr(Cmps)5   -0.287        -0.024        -0.025         0.113 -0.007
## fctr(NCls)1   -0.089        -0.127        -0.059         0.117 -0.140
## fctr(NCls)2   -0.152         0.023        -0.187         0.188 -0.025
## fctr(NCls)3    0.153        -0.020        -0.027         0.000  0.005
## fctr(NCls)4   -0.184         0.051        -0.255         0.052 -0.028
##               fc(C)3 fc(C)4 fc(C)5 f(NC)1 f(NC)2 f(NC)3
## fctr(PchL)1
## fctr(PchL)2
## fctr(PchL)3
## fctr(PchL)4
## fctr(PchL)5
## f(X19902000)2
## f(X19902000)3
## f(X19902000)4
## f(X19902000)5
## fctr(Cmps)1
## fctr(Cmps)2
## fctr(Cmps)3
## fctr(Cmps)4    0.011
## fctr(Cmps)5    0.013  0.205
## fctr(NCls)1    0.049 -0.004  0.183
```

```
## fctr(NCls)2   -0.405  0.028  0.047  0.166
## fctr(NCls)3    0.013 -0.523 -0.077  0.146  0.025
## fctr(NCls)4   -0.596  0.018  0.025  0.105  0.286  0.024
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
# Create indicator variables for PachListen, X1990s2000s, and Composing
for(i in 1:nrow(rate)){
  rate$pachind[i] <- ifelse(rate$PachListen[i] < 5, 0, 1)
}

for(i in 1:nrow(rate)){
  rate$x90sind[i] <- ifelse(rate$X1990s2000s[i] < 5, 0, 1)
}

for(i in 1:nrow(rate)){
  rate$compind[i] <- ifelse(rate$Composing[i] == 0, 0, 1)
}

# Run above model with indicator variables
mod16 <- lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subjec
t:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(mod16)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     (1 | Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8454.0   8527.3  -4214.0   8428.0     2058
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4744 -0.5485  0.0213  0.5221  3.4370
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.59318  0.7702
##  Subject:Voice      (Intercept) 0.06688  0.2586
##  Subject:Instrument (Intercept) 3.84923  1.9619
##  Residual                       2.36402  1.5375
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       5.06569    0.52568   9.637
## pachind           0.38828    0.47776   0.813
## x90sind           0.33888    0.35855   0.945
## compind           0.30421    0.35683   0.853
## factor(NoClass)1  0.07949    0.35149   0.226
## factor(NoClass)2 -0.17166    0.92456  -0.186
## factor(NoClass)3  1.01865    1.29351   0.788
## factor(NoClass)4 -1.45817    1.29268  -1.128
## factor(NoClass)8 -0.06749    0.92456  -0.073
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4
## pachind     -0.683
## x90sind     -0.502 -0.003
## compind     -0.181 -0.131  0.144
## fctr(NCls)1 -0.215 -0.197  0.001 -0.014
## fctr(NCls)2 -0.260  0.156  0.064 -0.084  0.176
## fctr(NCls)3  0.035 -0.055 -0.112 -0.194  0.164  0.054
## fctr(NCls)4  0.035 -0.055 -0.112 -0.194  0.164  0.054  0.091
## fctr(NCls)8 -0.260  0.156  0.064 -0.084  0.176  0.111  0.054  0.054
```



We looked at EDA of the variables to determine which covariates to try to add to our best model, the one with three random components. We tried to add the variables OMSI test score, familiarity with Pachelbel, listen to music from 1900s and 2000s, number of music classes, and level of composing. We added them one at a time and in groups. We compared the models on AIC and BIC. We found that our best model has an AIC of 8454 and a BIC of 8527.3. This model includes the 3 random components, a fixed effect for the number of classes, and fixed effects for three indicator variables we created: whether or not people fully recognized Pachelbel's Canon in D, whether or not people frequently listened to 90s and 2000s pop and rock, and if people have every composed before. Treating the variables in this way gave us the simplest model with the lowest information criteria, both of which were smaller than our orginal
5 model's information criteria.

Final list of variables: PachListen indicator, X1900s2000s indicator, Compsing indicator, NoClass, and random effects Subject:Instrument, Subject:Harmony, and Subject:Voice.

There's way too much detail in the work above. Also you have not found some fixed effects that seem to be important in the data.

b.

```
# Check if we should change any of the random effects
# Check voice
exactRLRT(lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subjec
t:Voice), data = rate), m0 = lmer(Classical ~ 1 + pachind + x90sind + compind + factor(No
Class) + (1|Subject:Instrument), data = rate), mA = lmer(Classical ~ 1 + pachind + x90sin
d + compind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Voice), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 0.77493, p-value = 0.1806
```

```
# Check harmony
exactRLRT(lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subjec
t:Harmony), data = rate), m0 = lmer(Classical ~ 1 + pachind + x90sind + compind + facto
r(NoClass) + (1|Subject:Instrument), data = rate), mA = lmer(Classical ~ 1 + pachind + x9
0sind + compind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmony), data =
rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 141.82, p-value < 2.2e-16
```

```
# Check instrument
exactRLRT(lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subjec
t:Instrument), data = rate), m0 = lmer(Classical ~ 1 + pachind + x90sind + compind + fact
or(NoClass) + (1|Subject:Harmony), data = rate), mA = lmer(Classical ~ 1 + pachind + x90s
ind + compind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmony), data = ra
te))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 1241.5, p-value < 2.2e-16
```

Now that we have included some fixed effects from other covariates, we will check to make sure the
random effects are still necessary. The p value for the REML likelihood ratio test on voice leading (the
9 MLE test cannot take this many random effects) is 0.177, which is lower than before but still not
significant. The p value for the random effect on harmony is 2.2x10^-16, so it is still significant and
should be included. The p value for instrument is the same as well, so will also continue to be included.

c.

```
summary(mod16)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     (1 | Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8454.0   8527.3  -4214.0   8428.0     2058
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4744 -0.5485  0.0213  0.5221  3.4370
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.59318  0.7702
##  Subject:Voice      (Intercept) 0.06688  0.2586
##  Subject:Instrument (Intercept) 3.84923  1.9619
##  Residual                       2.36402  1.5375
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       5.06569    0.52568   9.637
## pachind           0.38828    0.47776   0.813
## x90sind           0.33888    0.35855   0.945
## compind           0.30421    0.35683   0.853
## factor(NoClass)1  0.07949    0.35149   0.226
## factor(NoClass)2 -0.17166    0.92456  -0.186
## factor(NoClass)3  1.01865    1.29351   0.788
## factor(NoClass)4 -1.45817    1.29268  -1.128
## factor(NoClass)8 -0.06749    0.92456  -0.073
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4
## pachind     -0.683
## x90sind     -0.502 -0.003
## compind     -0.181 -0.131  0.144
## fctr(NCls)1 -0.215 -0.197  0.001 -0.014
## fctr(NCls)2 -0.260  0.156  0.064 -0.084  0.176
## fctr(NCls)3  0.035 -0.055 -0.112 -0.194  0.164  0.054
## fctr(NCls)4  0.035 -0.055 -0.112 -0.194  0.164  0.054  0.091
## fctr(NCls)8 -0.260  0.156  0.064 -0.084  0.176  0.111  0.054  0.054
```

From the model summary above, we can see the fit and effect each of the variables has on the model. To follow are the interpretations of our included fixed effects. For a participant who knows Pachelbel's Caonon in D, we expect to see an increase in classical ratings of 0.388 compared to those who don't, 9 holding all other variables constant. For a participant who listens to pop and rock from the 90s and 2000s, we expect to see an increas in classical ratings of 0.339 compared to those who don't, holding all other variables constant. For someone who has experience composing, we expect to see an increase in classical ratings of 0.304 compared to those with no experience, holding all other variables constant. For those who have taken 1, 2, 3, 4, or 8 music classes, we expect to see a change in classical ratings of 0.079, -0.172, 1.019, -1.458, and -0.067 respectively, compared to participants who have never taken a music class, holding all other variables constant. We can also look at the correlations between the fixed effects, which the largest being -0.197 between taking one music class and knowing Pachelbel's Canon in D. We also have the random effects for voice leading, harmony, and instrument, which have variances of 0.067, 0.593, and 3.849, respectively. The overall model AIC is 8454 and the BIC is 8527.3.

3.

```
# Create indicator variable for self declared musician or not
# Split based on making groups two equal sizes
for(i in 1:nrow(rate)){
  rate$musicind[i] <- ifelse(rate$Selfdeclare[i] <=2, 0, 1)
}

# Add self decalred musician indicator to the model as a fixed effect
mod17 <- lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + musicind +
(1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML = FAL
SE)
summary(mod17)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     musicind + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8455.4   8534.3  -4213.7   8427.4     2057
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4722 -0.5491  0.0200  0.5200  3.4407
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.59282  0.7699
##  Subject:Voice      (Intercept) 0.06678  0.2584
##  Subject:Instrument (Intercept) 3.83477  1.9583
##  Residual                       2.36409  1.5376
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       5.15118    0.53519   9.625
## pachind           0.37841    0.47714   0.793
## x90sind           0.30320    0.36063   0.841
## compind           0.44556    0.39618   1.125
## factor(NoClass)1  0.10969    0.35286   0.311
## factor(NoClass)2 -0.14714    0.92355  -0.159
## factor(NoClass)3  1.15317    1.30190   0.886
## factor(NoClass)4 -1.32366    1.30107  -1.017
## factor(NoClass)8 -0.04297    0.92355  -0.047
## musicind         -0.31581    0.38730  -0.815
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4 f(NC)8
## pachind     -0.674
## x90sind     -0.512  0.000
## compind     -0.074 -0.129  0.076
## fctr(NCls)1 -0.190 -0.198 -0.012  0.033
## fctr(NCls)2 -0.248  0.155  0.060 -0.061  0.178
## fctr(NCls)3  0.059 -0.058 -0.126 -0.117  0.175  0.057
## fctr(NCls)4  0.059 -0.058 -0.126 -0.118  0.175  0.057  0.105
## fctr(NCls)8 -0.248  0.155  0.060 -0.061  0.178  0.112  0.057  0.057
## musicind    -0.196  0.025  0.121 -0.438 -0.105 -0.033 -0.127 -0.127 -0.033
```

Adding the self declared musician indicator variable to the model, we can see that it changes to effects of some of the other variables. For examples, it decreased the effect of having taken 8 music classes, and some of this effect is now being credited to being a musician. This is probably because these variables are correlated, as people who have taken 8 music classes would probably consider themselves to be musicians. The AIC and BIC are higher for this model, so we do not think that this variable needs to be added in this way.

```
# Add musican indicator as interaction with composer indicator
mod18 <- lmer(Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) + compind*mus
icind + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, RE
ML = FALSE)
summary(mod18)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Classical ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     compind * musicind + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8453.5   8538.1  -4211.8   8423.5     2056
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4644 -0.5493  0.0153  0.5191  3.4526
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.59069  0.7686
##  Subject:Voice      (Intercept) 0.06627  0.2574
##  Subject:Instrument (Intercept) 3.75279  1.9372
##  Residual                       2.36452  1.5377
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)         4.9176     0.5435   9.049
## pachind             0.3848     0.4727   0.814
## x90sind             0.2917     0.3573   0.816
## compind             1.2970     0.5857   2.215
## factor(NoClass)1    0.3075     0.3639   0.845
## factor(NoClass)2   -0.6608     0.9518  -0.694
## factor(NoClass)3    1.5813     1.3082   1.209
## factor(NoClass)4   -0.8955     1.3074  -0.685
## factor(NoClass)8    0.2878     0.9304   0.309
## musicind            0.3322     0.5066   0.656
## compind:musicind   -1.6889     0.8623  -1.959
##
## Correlation of Fixed Effects:
##              (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4 f(NC)8
## pachind     -0.659
## x90sind     -0.496  0.000
## compind     -0.211 -0.081  0.038
## fctr(NCls)1 -0.239 -0.188 -0.016  0.227
## fctr(NCls)2 -0.172  0.147  0.062 -0.244  0.088
## fctr(NCls)3  0.020 -0.056 -0.127  0.046  0.212  0.008
## fctr(NCls)4  0.020 -0.056 -0.127  0.046  0.212  0.008  0.130
## fctr(NCls)8 -0.278  0.153  0.056  0.094  0.219  0.056  0.086  0.086
## musicind    -0.288  0.024  0.081  0.263  0.105 -0.204  0.014  0.014  0.094
## cmpnd:mscnd  0.219 -0.007  0.016 -0.742 -0.278  0.276 -0.167 -0.167 -0.182
##              muscnd
## pachind
```

```
## x90sind
## compind
## fctr(NCls)1
## fctr(NCls)2
## fctr(NCls)3
## fctr(NCls)4
## fctr(NCls)8
## musicind
## cmpnd:mscnd -0.653
```

7

We then looked to see if adding this variable as an interaction with another variable it is slightly correlated with would help the model. Adding the musican indicator as an interaction with the composer indicator also did not help our model much in terms of AIC or BIC, but the interacton term is significant with a t statistic of -1.959. Because of this, we will keep it in our model.

what about interactions of musicind with other fixed effects?

4.

a.

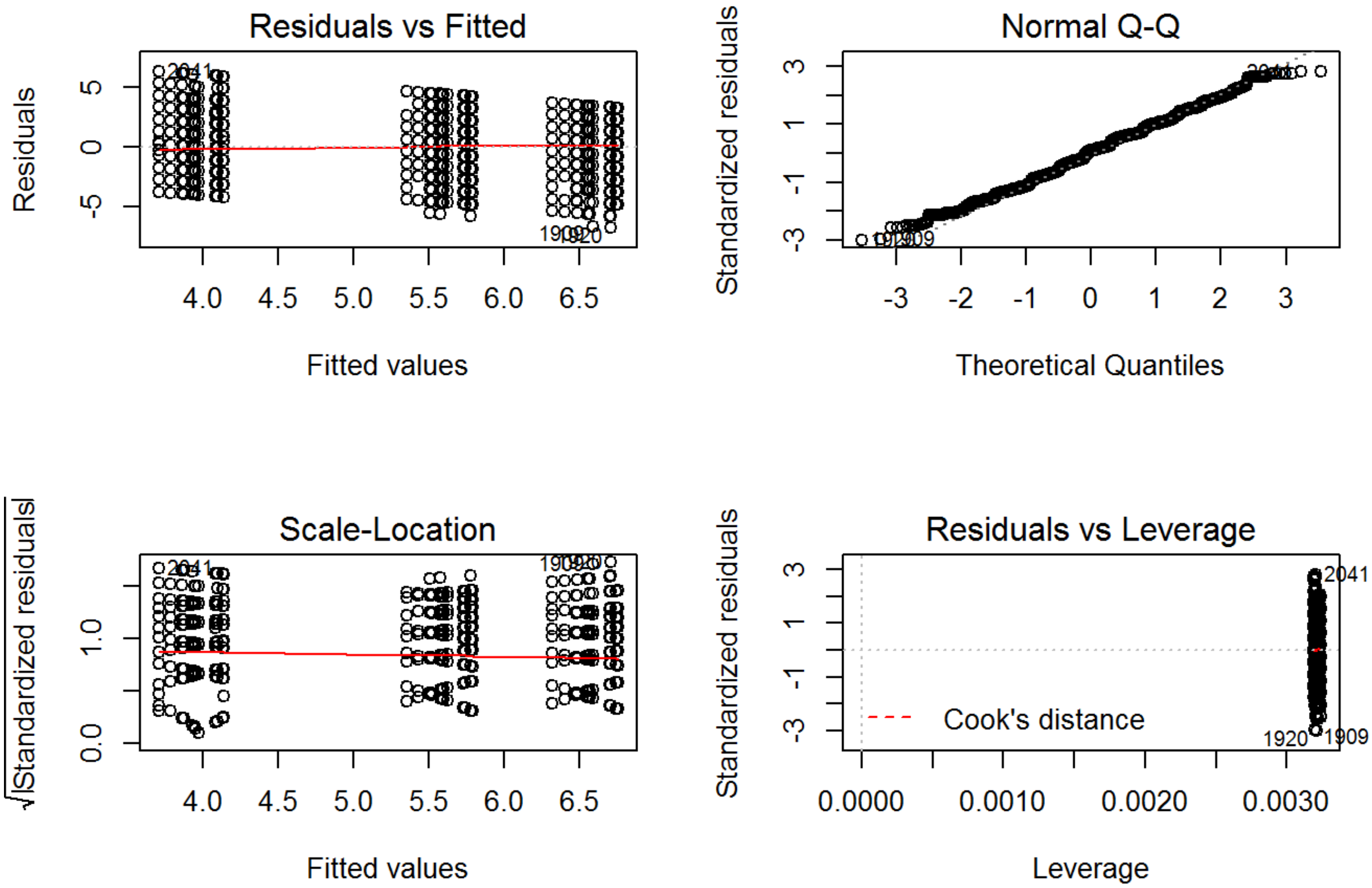
```
# MAKE NEW MODELS WITH POPULAR AS RESPONSE

# Fit linear regression to predict popular rating from instrument, harmony, and voice
pop1 <- lm(Popular ~ Instrument + Harmony + Voice, data = rate)
summary(pop1)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7130 -1.7024  0.2099  1.4782  6.2903
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.59238    0.12677  52.004   <2e-16 ***
## Instrumentpiano  -0.96819    0.11030  -8.777   <2e-16 ***
## Instrumentstring -2.61435    0.10964 -23.845   <2e-16 ***
## HarmonyI-V-IV    -0.04538    0.12701  -0.357   0.7209
## HarmonyI-V-VI    -0.26837    0.12696  -2.114   0.0346 *
## HarmonyIV-I-V    -0.18924    0.12690  -1.491   0.1360
## Voicepar3rd       0.15535    0.11007   1.411   0.1583
## Voicepar5th       0.16595    0.10997   1.509   0.1314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.242 on 2483 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.192,  Adjusted R-squared:  0.1897
## F-statistic:  84.3 on 7 and 2483 DF,  p-value: < 2.2e-16
```

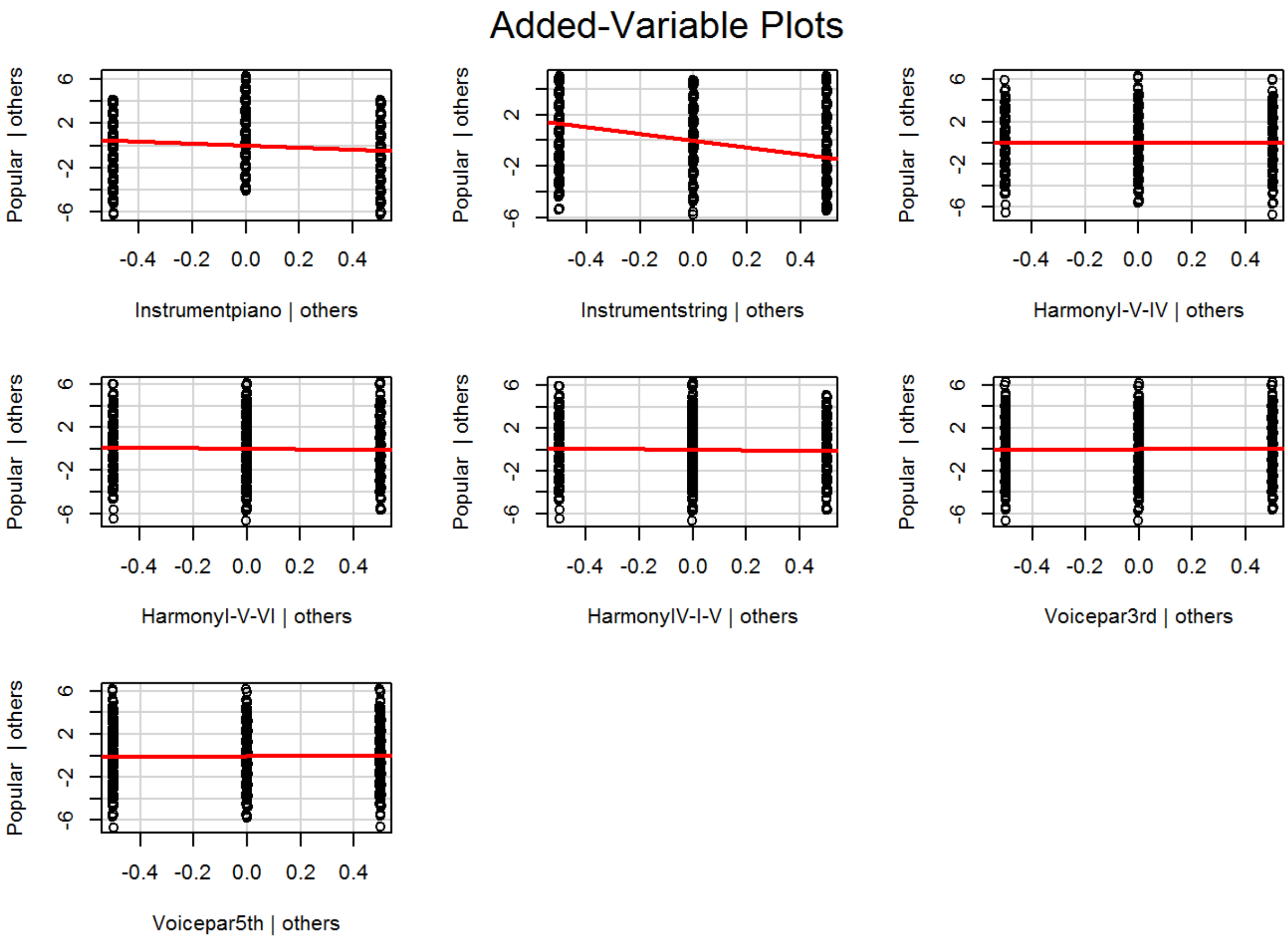
```
AIC(pop1)
```

```
## [1] 11100.47
```

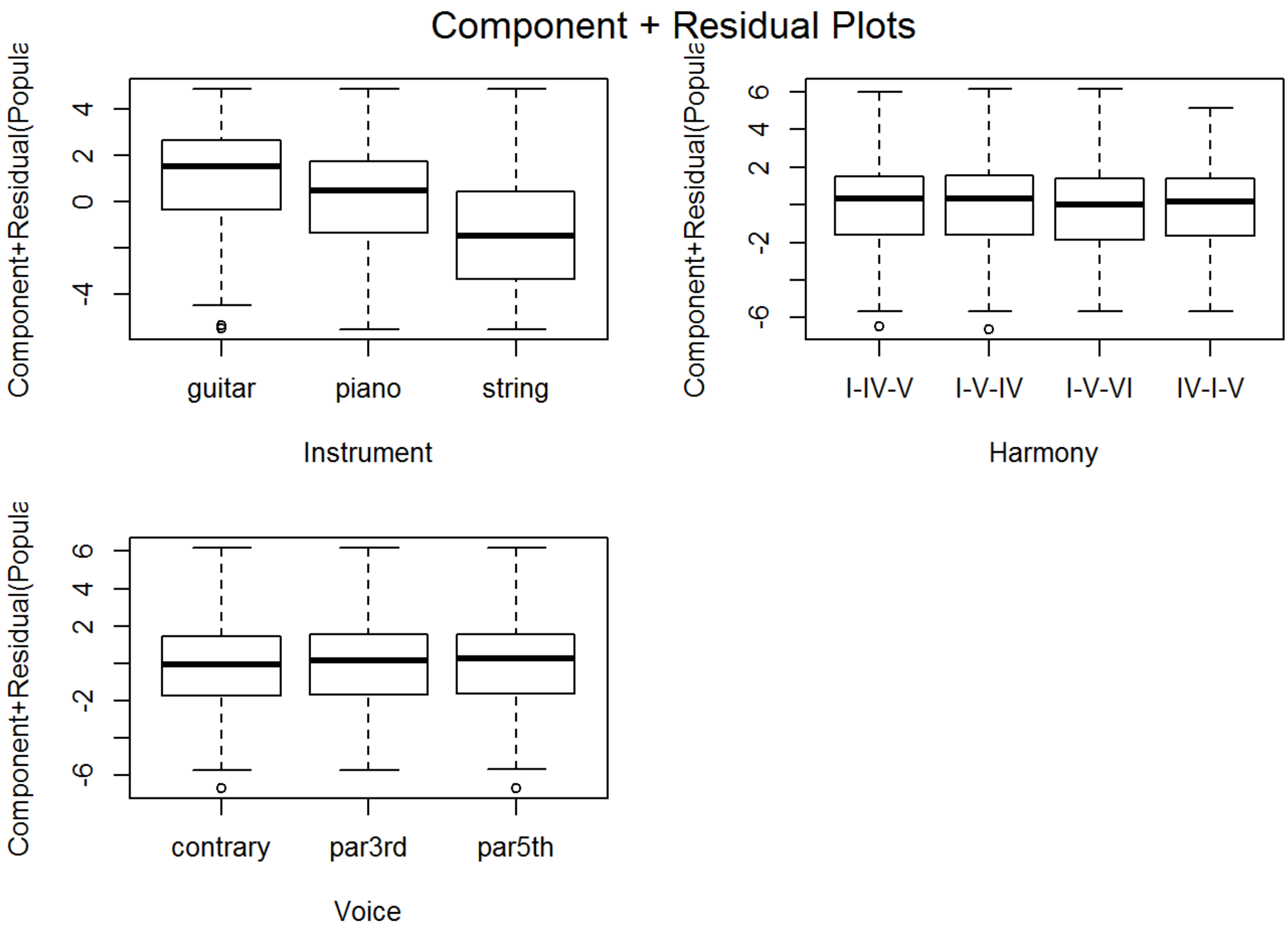
```
# Check model diagnostics
par(mfrow=c(2,2))
plot(pop1)
```




```
avPlots(pop1)
```

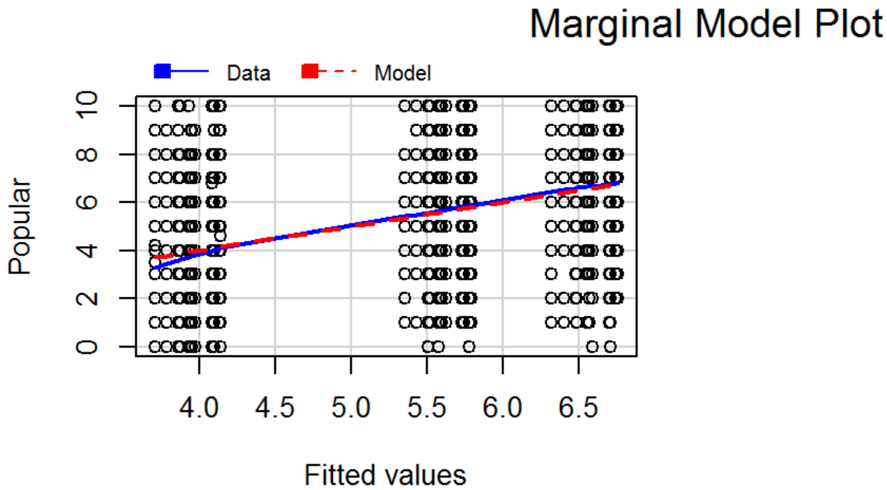



```
crPlots(pop1)
```




```
marginalModelPlots(pop1)
```

```
## Warning in mmps(...): Interactions and/or factors skipped
```



We will now refit many of our models with popular ratings as the response. We started with a linear regression with all three key variables. We can see from the summary statistics above that instrument is very important, as well as one type of harmony. Voice leading is not significant at any level. The AIC is 11100.47 and the adjusted R squared is 0.1897. We will compare these values to those of models which each of the three variables to deterine if they are actually influential and need to be included in the model. We also checked the model diagnostics to make sure that the linear model assumptions are met so that we can report on the p values and significance of the coefficients. The residuals have a small trend, but the QQ plot and leverage values lookk sufficient to fulfill our linear model assumptions. The added variable plots, component + residual plots, and marginal model plots all look as we would expect.

```
# Fit linear regression to predict popular ratings from harmony and voice
pop2 <- lm(Popular ~ Harmony + Voice, data = rate)
summary(pop2)
```

```
##
## Call:
## lm(formula = Popular ~ Harmony + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5610 -1.5610 -0.1254  1.7226  4.8746
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.39647    0.12196  44.248   <2e-16 ***
## HarmonyI-V-IV -0.04841    0.14104  -0.343   0.7315
## HarmonyI-V-VI -0.27104    0.14098  -1.922   0.0547 .
## HarmonyIV-I-V -0.18889    0.14093  -1.340   0.1802
## Voicepar3rd    0.15192    0.12223   1.243   0.2140
## Voicepar5th    0.16453    0.12212   1.347   0.1780
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.489 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.002792,   Adjusted R-squared:  0.0007853
## F-statistic: 1.391 on 5 and 2485 DF,  p-value: 0.2243
```

```
AIC(pop2)
```

```
## [1] 11620.63
```

```
# Fit linear regression to predict popular ratings from instrument and voice
pop3 <- lm(Popular ~ Instrument + Voice, data = rate)
summary(pop3)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Voice, data = rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6321 -1.6321  0.1476  1.5015  6.1476
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.4668     0.1004  64.432   <2e-16 ***
## Instrumentpiano   -0.9683     0.1104  -8.773   <2e-16 ***
## Instrumentstring  -2.6144     0.1097 -23.832   <2e-16 ***
## Voicepar3rd        0.1559     0.1101   1.415    0.157
## Voicepar5th        0.1653     0.1100   1.503    0.133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.243 on 2486 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.1901, Adjusted R-squared:  0.1888
## F-statistic: 145.9 on 4 and 2486 DF,  p-value: < 2.2e-16
```

```
AIC(pop3)
```

```
## [1] 11100.27
```

```
# Fit linear regression to predict popular ratings from instrument and harmony
pop4 <- lm(Popular ~ Instrument + Harmony, data = rate)
summary(pop4)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony, data = rate)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.699 -1.686  0.183  1.490  6.183
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.69921    0.10981  61.007   <2e-16 ***
## Instrumentpiano  -0.96812    0.11032  -8.775   <2e-16 ***
## Instrumentstring -2.61416    0.10966 -23.839   <2e-16 ***
## HarmonyI-V-IV    -0.04494    0.12703  -0.354   0.7235
## HarmonyI-V-VI    -0.26807    0.12698  -2.111   0.0349 *
## HarmonyIV-I-V    -0.18911    0.12693  -1.490   0.1364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.242 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.1911, Adjusted R-squared:  0.1895
## F-statistic: 117.4 on 5 and 2485 DF,  p-value: < 2.2e-16
```

```
AIC(pop4)
```

```
## [1] 11099.33
```

We also refit our linear model removing one predictor at a time to see if that helped our model in terms of adjusted R squared and AIC. For the model without instrument, the AIC increased and adjusted R squared decreased significantly, meaning that instrument was helping us predict popular ratings. This is to be expected since it was significant in our original model. The AIC and adjusted R squared are about the same as the original model for our linear model without hamrony. For out model without voice leading, the AIC and adjusted R sqared are also about the same. However, we will continue to include them because they are key experimental factors.

```
# Fit hierarchical model to predict popular ratings with fixed effects for 3 variables an
d random intercept
pop5 <- lmer(Popular ~ Instrument + Harmony + Voice + (1|Subject), data = rate, REML = FA
LSE)
exactRLRT(pop5)
```

```
## Using restricted likelihood evaluated at ML estimators.
## Refit with method="REML" for exact results.
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 712.12, p-value < 2.2e-16
```

```
summary(pop5)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##  10390.2  10448.4  -5185.1  10370.2     2481
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7477 -0.6477  0.0400  0.6684  3.1832
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.520    1.233
##  Residual             3.478    1.865
## Number of obs: 2491, groups:  Subject, 70
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        6.58561    0.18120   36.34
## Instrumentpiano   -0.95930    0.09186  -10.44
## Instrumentstring  -2.60902    0.09125  -28.59
## HarmonyI-V-IV     -0.04241    0.10567   -0.40
## HarmonyI-V-VI     -0.27232    0.10563   -2.58
## HarmonyIV-I-V     -0.18932    0.10558   -1.79
## Voicepar3rd        0.15980    0.09158    1.75
## Voicepar5th        0.16760    0.09149    1.83
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.251
## Instrmntstr -0.252  0.498
## HrmnyI-V-IV -0.290  0.002    -0.001
## HrmnyI-V-VI -0.290  0.001    -0.001     0.499
## HrmnyIV-I-V -0.291 -0.001     0.000     0.499  0.499
## Voicepar3rd -0.252  0.000    -0.001    -0.001  0.001  0.001
## Voicepar5th -0.252 -0.001    -0.001    -0.002 -0.003 -0.002  0.500
```

Our next model fit to predict popular ratings is a hierarchical model with a random intercept for each participant with fixed effects for each of the key variables. From the REML LRT, we can see that there is an extremely small p value, menaing that we should keep the random effect in this model with the other variables. We also looked at the model summary, where we can see the AIC of 10390.2 and the coefficients on the fixed effects, many of which are significant in this model.

```
# Fit model with 3 variables as random effects
pop6 <- lmer(Popular ~ 1 + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voic
e), data = rate, REML = FALSE)
summary(pop6)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ 1 + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##  10124.7  10153.8  -5057.3  10114.7     2486
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6744 -0.5770 -0.0111  0.5774  3.1739
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.37175  0.6097
##  Subject:Voice      (Intercept) 0.02728  0.1652
##  Subject:Instrument (Intercept) 3.25149  1.8032
##  Residual                       2.47271  1.5725
## Number of obs: 2491, groups:
## Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    5.379      0.134   40.15
```

```
# Compare linear regression, random intercept model with fixed effects, and model with 3
ranodm components
anova(pop6, pop5, pop1)
```

```
## Data: rate
## Models:
## pop6: Popular ~ 1 + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
## pop6:     (1 | Subject:Voice)
## pop1: Popular ~ Instrument + Harmony + Voice
## pop5: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
##      Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## pop6  5 10125 10154 -5057.3    10115
## pop1  9 11100 11153 -5541.2    11082   0.00      4          1
## pop5 10 10390 10448 -5185.1    10370 712.23      1     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finally, we fit a model with random intercepts for each participant and factor of the key variables combination. The AIC is a bit lower at 10124.7, and instrument has a large variance in the model, along with the residuals. For comparison and AIC purposes, we will use this model going forward in our analysis of popular rationgs. Instrument has the largest effect on popular ratings when there are only
9 these 3 variables in the data. This can be seen from the significant and size of variance in each of the models. The next most important is harmony, and then voice leading, again for the same reasons.

b.

```
# Fit model with other covariates
pop7 <- lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subject:Ins
trument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML = FALSE)
summary(pop7)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     (1 | Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8449.2   8522.5  -4211.6   8423.2     2058
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6669 -0.5768  0.0029  0.5776  3.1658
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.39985  0.6323
##  Subject:Voice      (Intercept) 0.03202  0.1789
##  Subject:Instrument (Intercept) 3.17333  1.7814
##  Residual                       2.48484  1.5763
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       5.16199    0.47442  10.881
## pachind          -0.33264    0.43117  -0.771
## x90sind           0.36342    0.32361   1.123
## compind           0.57503    0.32210   1.785
## factor(NoClass)1  0.01164    0.31725   0.037
## factor(NoClass)2 -0.58295    0.83441  -0.699
## factor(NoClass)3  2.19319    1.16760   1.878
## factor(NoClass)4 -0.15668    1.16665  -0.134
## factor(NoClass)8  0.03511    0.83441   0.042
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4
## pachind     -0.683
## x90sind     -0.502 -0.003
## compind     -0.182 -0.131  0.144
## fctr(NCls)1 -0.215 -0.197  0.000 -0.015
## fctr(NCls)2 -0.260  0.156  0.064 -0.084  0.176
## fctr(NCls)3  0.035 -0.055 -0.112 -0.194  0.164  0.054
## fctr(NCls)4  0.035 -0.055 -0.112 -0.194  0.164  0.054  0.091
## fctr(NCls)8 -0.260  0.156  0.064 -0.084  0.176  0.111  0.054  0.054
```

We added the same covariates from the classical model to our popular model so that they will be comparable. The model including these new covariates has an AIC of 8449.2 and a BIC of 8522.2, which are very similar to the statistics we got from the classicial models, so they are fitting the different variables similarly. This makes sense since the distributions of classical and popular ratings are very similar.

```
# Check if we should change any of the random effects
# Check voice
exactRLRT(lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subject:V
oice), data = rate), m0 = lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClas
s) + (1|Subject:Instrument), data = rate), mA = lmer(Popular ~ 1 + pachind + x90sind + co
mpind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Voice), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 0.049392, p-value = 0.3958
```

```
# Check harmony
exactRLRT(lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subject:H
armony), data = rate), m0 = lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoCla
ss) + (1|Subject:Instrument), data = rate), mA = lmer(Popular ~ 1 + pachind + x90sind + c
ompind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmony), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 76.327, p-value < 2.2e-16
```

```
# Check instrument
exactRLRT(lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + (1|Subject:I
nstrument), data = rate), m0 = lmer(Popular ~ 1 + pachind + x90sind + compind + factor(No
Class) + (1|Subject:Harmony), data = rate), mA = lmer(Popular ~ 1 + pachind + x90sind + c
ompind + factor(NoClass) + (1|Subject:Instrument) + (1|Subject:Harmony), data = rate))
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 993.6, p-value < 2.2e-16
```

```
summary(pop7)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     (1 | Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8449.2   8522.5  -4211.6   8423.2     2058
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6669 -0.5768  0.0029  0.5776  3.1658
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.39985  0.6323
##  Subject:Voice      (Intercept) 0.03202  0.1789
##  Subject:Instrument (Intercept) 3.17333  1.7814
##  Residual                       2.48484  1.5763
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       5.16199    0.47442  10.881
## pachind          -0.33264    0.43117  -0.771
## x90sind           0.36342    0.32361   1.123
## compind           0.57503    0.32210   1.785
## factor(NoClass)1  0.01164    0.31725   0.037
## factor(NoClass)2 -0.58295    0.83441  -0.699
## factor(NoClass)3  2.19319    1.16760   1.878
## factor(NoClass)4 -0.15668    1.16665  -0.134
## factor(NoClass)8  0.03511    0.83441   0.042
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4
## pachind     -0.683
## x90sind     -0.502 -0.003
## compind     -0.182 -0.131  0.144
## fctr(NCls)1 -0.215 -0.197  0.000 -0.015
## fctr(NCls)2 -0.260  0.156  0.064 -0.084  0.176
## fctr(NCls)3  0.035 -0.055 -0.112 -0.194  0.164  0.054
## fctr(NCls)4  0.035 -0.055 -0.112 -0.194  0.164  0.054  0.091
## fctr(NCls)8 -0.260  0.156  0.064 -0.084  0.176  0.111  0.054  0.054
```

GIven the fixed effects that we just added, we will check to make sure the random effects are still important. The p value for the REML likelihood ratio test on voice leading is 0.4076, which is not significant. The p value for harmony is 2.2x10^-16, as it is for instrument. This tells us that harmony and

9 instrument are very significant in predicting popular ratings. This is the same result we got in previous models, an so for the classical models, so this is a reassuring sign. The fixed effects on composer and taking 3 music classes are also significant.

c.

```
# Add self decalred musician indicator to the model as a fixed effect
pop8 <- lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + musicind +
(1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML = FAL
SE)
summary(pop8)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     musicind + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8451.2   8530.1  -4211.6   8423.2     2057
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6669 -0.5768  0.0029  0.5776  3.1658
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.39985  0.6323
##  Subject:Voice      (Intercept) 0.03202  0.1789
##  Subject:Instrument (Intercept) 3.17333  1.7814
##  Residual                       2.48484  1.5763
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)       5.162036   0.483800  10.670
## pachind          -0.332643   0.431312  -0.771
## x90sind           0.363398   0.326023   1.115
## compind           0.575104   0.358160   1.606
## factor(NoClass)1  0.011654   0.319016   0.037
## factor(NoClass)2 -0.582933   0.834849  -0.698
## factor(NoClass)3  2.193262   1.177105   1.863
## factor(NoClass)4 -0.156608   1.176161  -0.133
## factor(NoClass)8  0.035123   0.834849   0.042
## musicind         -0.000176   0.350125  -0.001
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4 f(NC)8
## pachind     -0.674
## x90sind     -0.512  0.000
## compind     -0.074 -0.129  0.076
## fctr(NCls)1 -0.189 -0.198 -0.012  0.033
## fctr(NCls)2 -0.248  0.155  0.060 -0.061  0.178
## fctr(NCls)3  0.059 -0.058 -0.126 -0.118  0.175  0.057
## fctr(NCls)4  0.059 -0.058 -0.126 -0.118  0.175  0.057  0.105
## fctr(NCls)8 -0.248  0.155  0.060 -0.061  0.178  0.112  0.057  0.057
## musicind    -0.196  0.025  0.121 -0.437 -0.105 -0.033 -0.127 -0.127 -0.033
```

Adding a fixed effect for musician indicator did not change the variances on any of the random effects. It also barely changed the coefficients on any of the fixed effects either. Since it is not significant, we conclude that we do not need this variable included in this way.

```
# Add musican indicator as interaction with composer indicator
pop9 <- lmer(Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) + compind*musici
nd + (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice), data = rate, REML
= FALSE)
summary(pop9)
```

comment similar to comment on #3

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Popular ~ 1 + pachind + x90sind + compind + factor(NoClass) +
##     compind * musicind + (1 | Subject:Instrument) + (1 | Subject:Harmony) +
##     (1 | Subject:Voice)
##    Data: rate
##
##      AIC      BIC   logLik deviance df.resid
##   8450.8   8535.4  -4210.4   8420.8     2056
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6664 -0.5769 -0.0005  0.5755  3.1590
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  Subject:Harmony    (Intercept) 0.39900  0.6317
##  Subject:Voice      (Intercept) 0.03177  0.1782
##  Subject:Instrument (Intercept) 3.12950  1.7690
##  Residual                       2.48513  1.5764
## Number of obs: 2071, groups:
## Subject:Harmony, 232; Subject:Voice, 174; Subject:Instrument, 174
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        5.32807    0.49288  10.810
## pachind           -0.33713    0.42872  -0.786
## x90sind            0.37171    0.32410   1.147
## compind           -0.03051    0.53117  -0.057
## factor(NoClass)1 -0.12918    0.33008  -0.391
## factor(NoClass)2 -0.21763    0.86321  -0.252
## factor(NoClass)3  1.88832    1.18672   1.591
## factor(NoClass)4 -0.46150    1.18579  -0.389
## factor(NoClass)8 -0.20034    0.84385  -0.237
## musicind          -0.46106    0.45947  -1.003
## compind:musicind  1.20154    0.78211   1.536
##
## Correlation of Fixed Effects:
##             (Intr) pachnd x90snd compnd f(NC)1 f(NC)2 f(NC)3 f(NC)4 f(NC)8
## pachind     -0.659
## x90sind     -0.496  0.000
## compind     -0.211 -0.081  0.038
## fctr(NCls)1 -0.238 -0.189 -0.017  0.227
## fctr(NCls)2 -0.172  0.147  0.062 -0.244  0.088
## fctr(NCls)3  0.020 -0.056 -0.127  0.046  0.212  0.008
## fctr(NCls)4  0.020 -0.056 -0.127  0.046  0.213  0.008  0.130
## fctr(NCls)8 -0.278  0.153  0.056  0.094  0.219  0.056  0.086  0.086
## musicind    -0.288  0.024  0.081  0.263  0.105 -0.204  0.014  0.015  0.094
## cmpnd:mscnd  0.219 -0.007  0.017 -0.742 -0.278  0.275 -0.167 -0.167 -0.182
##             muscnd
## pachind
```

```
## x90sind
## compind
## fctr(NCls)1
## fctr(NCls)2
## fctr(NCls)3
## fctr(NCls)4
## fctr(NCls)8
## musicind
## cmpnd:mscnd -0.653
```

We again added the musician indicator as an interaction with composer. This was a much more positive result and has a larger impact on our ability to confidently predict popular ratings. The AIC of this new model is 8450.8 and the BIC is 8535.4. Adding this terms decreased the variance on the instrument random effect, probably because musician can match instruments to music types better.

5.

We are interested in measuring which covariates are best at predicting how classical and popular people will ranks songs that they hear. Specifically, we are interested in learning if instrument, harmonic motion, and voice leading have a strong impact on what participants in the study rate the songs. It is hypothesized that instrument will have the largest effect of the three key variables, and certain harmonies and voice leadings, I-V-vi and contrary motion, respectively, will be frequently rated as classical. We have data on 2520 song ratings, for 70 study participants. We built hierarchical linear models to find answers to our hypotheses.

First, we built models to predict classical ratings. In part 2, we found that instrument and harmony are very significant in our model, and instrument especially contributes a lot to explaining the variation in the data. The most predictive model we built had three random components for each of the key variables, and also fixed effects for an indicator if people know Pachelbel's Canon in D, an indicator if people listened to pop and rock from the 1990s and 2000s, an indicator if people have experience composing, how many music classes they have taken, an indicator if they are a self-declared musician, and the interaction between being a musician and a composer. Individually, most of these variables were not significant in our final model, but we kept them in because they were significant earlier, and when removed, measures such as AIC and BIC increased.

In order to predict what popular ratings people gave songs, we built many of the same models as before.
It ended up being the case that our final models were the same for each case (with different response
10 variables). The model also had the lowest AIC and BIC, even though similarly, many of the variables
were not significant on their own. This is not a standard repeated measures model, as it takes random draws for every participant-instrument, participant-harmony, and participant-voice combination, which adds an extra aspect of variance to the model.

Overall, in order to address Dr. Jimenez's original hypotheses, we can confirm that he was correct in assuming that instrument has the largest effect on ratings. No other variable, in the three experimental variables or in others we added, had as much predictive power as instrument. This can be seen from each of the models we built, from the linear regression to the model with random effects, to the hierarchical models with other covariates and interactions. From our linear model, we can see that harmonic motion I-V-vi is very predictive of classical ratings, as it has a small p value. It is also very significant in the model with harmony as a fixed effect. From those same two models mentioned with respect to harmonic motion, we can also determine that contrary motion is frequently rated as classical because the coefficients on the other two types of voice leading are negative, so they are less likely to be classical than contrary motion. These two coefficients are significant in both models, so at the 95% confidence level we can be confident in their implications. In the end, we have decided on one model each to predict classical and popular ratings of songs, using a variety of ways to integrate the variables such that the AIC and BIC are minimized.