Stephanie Gill

36-763

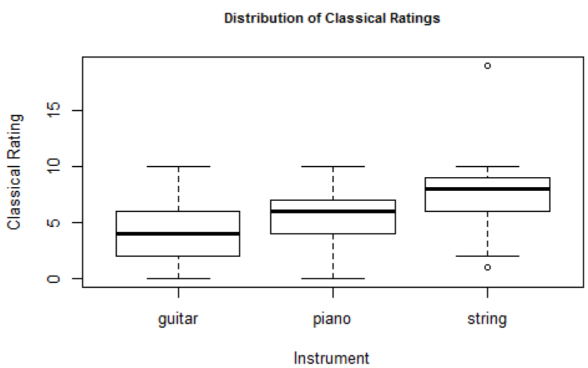
December 10th, 2013

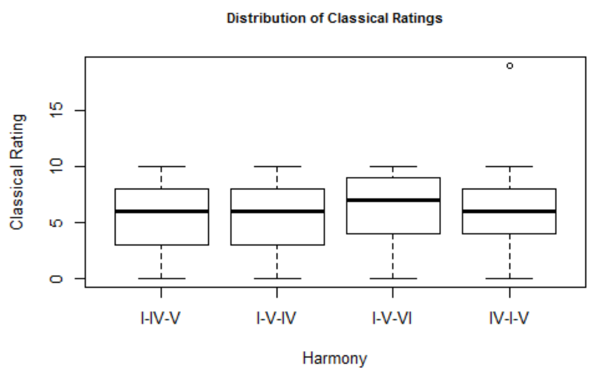
Homework 5

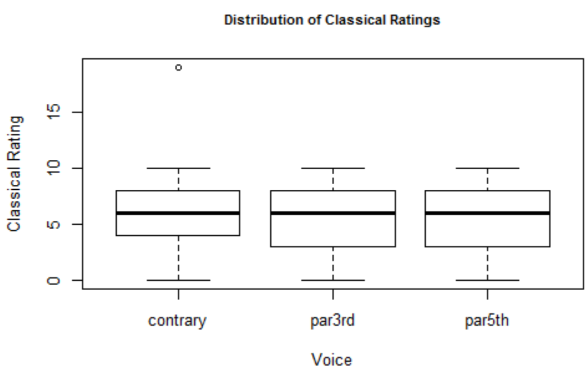
Exercises

1.

a. First, I made some boxplots to observe the distribution of classical ratings across different instruments, harmonies, and voices.







There does not seem to be any significant differences in the distribution of classical ratings between each categories of each of the three predictor variables. However, in the first plot, we do see that string quartets have a higher mean classical rating than the other two instrument categories. The second plot shows that I-V-VI harmony has the highest median classical rating. The third plot shows no noticeable differences in median classical rating across the three groups. While further exploring the variables, I discovered that there are NA values in the dataset. For example, Classical and Popular, which are the two response variables in my analysis, are both missing the same 27 values. Therefore, I subset the dataset to exclude rows with missing values for the response.

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

attach(ratings2)
```

Then, the conventional linear model using the lm function is the following:

```
fit.lm.full <- lm(Classical~ Instrument + Harmony + Voice)
```

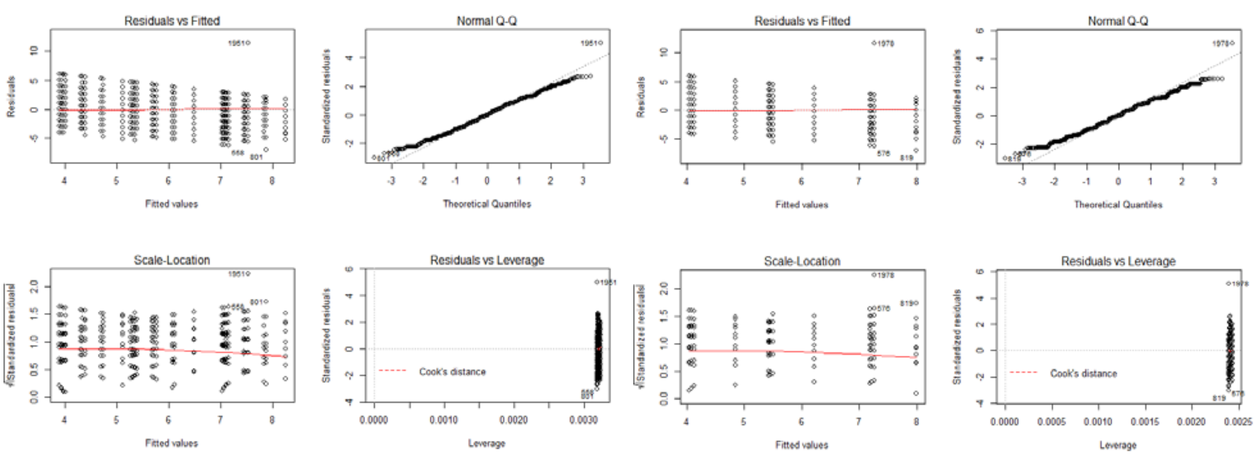
Then, I also made conventional linear model with classical as the response variable, and taking out each variable each time.

```
fit.lm.1 <- lm(Classical~ Harmony + Voice)

fit.lm.2 <- lm(Classical~Instrument+Voice)

fit.lm.3 <- lm(Classical~Instrument+Harmony)

fit.lm.4 <- lm(Classical~Instrument)

fit.lm.5 <- lm(Classical~Harmony)

fit.lm.6 <- lm(Classical~Voice)
```

I was then able to compare these 7 models using AIC and BIC.

	Fit.lm.full	Fit.lm.1	Fit.lm.2	Fit.lm.3	Fit.lm.4	Fit.lm.5	Fit.lm.6
AIC	11230.45	11908.94	11275.96	11242.96	11287.86	11917.23	11942.32
BIC	11282.84	11949.69	11310.89	11283.43	11311.14	11946.34	11965.61

Looking at AIC, the best model is the full model, which includes all three predictors. Looking at BIC, we are indifferent between the full model and the third model, which is a model including instrument and harmony. However, we must keep in mind that BIC prefers simpler models and penalizes adding a predictor more severely than AIC.



Figures: Diagnostic plots for fit.lm.full (left) and for fit.lm.3 (right)

Looking at the diagnostic plots, it is hard to tell which model is fitting better. However, adjusted R^2 values are 0.2529 and 0.2487 for fit.lm.full and fit.lm.3. Moving forward, I will choose the full model since the AIC and adjusted R^2 both prefer this model over the third model excluding voice. The resulting full model has the following output:

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.34016    0.12987  33.420  < 2e-16 ***
Instrumentpiano     1.37359    0.11298  12.158  < 2e-16 ***
Instrumentstring    3.13312    0.11230  27.899  < 2e-16 ***
HarmonyI-V-IV      -0.03108    0.13008  -0.239 0.811168
HarmonyI-V-VI       0.76909    0.13008   5.913 3.83e-09 ***
HarmonyIV-I-V       0.05007    0.12997   0.385 0.700092
Voicepar3rd        -0.41247    0.11271  -3.660 0.000258 ***
Voicepar5th        -0.37058    0.11264  -3.290 0.001016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 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
```

According to the output, all three predictors are statistically significant. Therefore, I conclude that Harmony, Instrument, and Voice are all influential predictors on classical ratings. Furthermore, both stimuli with piano and stimuli with string quartets have higher classical ratings than stimuli with electric guitar. Similarly, harmony I-V-VI has a higher classical rating on average compared to Harmony I-IV-V. Lastly, both stimuli with voice par3rd and voice par5th have lower classical rating on average compared to voice contrary.

b.

i. Mathematical Model:

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

ii.

```
M1 <- lmer(Classical ~ Instrument + Harmony + Voice + (1 | Subject) )
```

The corresponding model is noted above. This model includes a random intercept for each of the 70 subjects for the three fixed effects. In this part, we are asked to assess whether or not this random intercept is needed in the model. In order to do this, we can see whether or not this model does better than the conventional linear model from part a, including all three predictors.

	M1	Fit.lm.full
AIC	10491.51	11230.45
BIC	10549.73	11282.84

AIC and BIC both prefer the model with the random intercept by a fairly large amount. Meaning, we can conclude that this random intercept is needed in the model.

Another method to test if the random intercept is needed is to compare the residuals for the two models graphically. The residual plot for fit.lm.full which is the linear regression model including no random effect is shown below. There is a trend in the plot—Residuals seem to slope down and are not spread out evenly around the horizontal line marked in red. The three sets of residual plots for M1, the model including a random intercept, is also shown below. For each subject, conditional and marginal residual plots look good. Meaning, the residual values seem to be evenly spread out around the horizontal lines marked in blue and we observe no clear trend or patterns. This is further evidence that the random effects model performs better than the conventional linear model with no random effects.

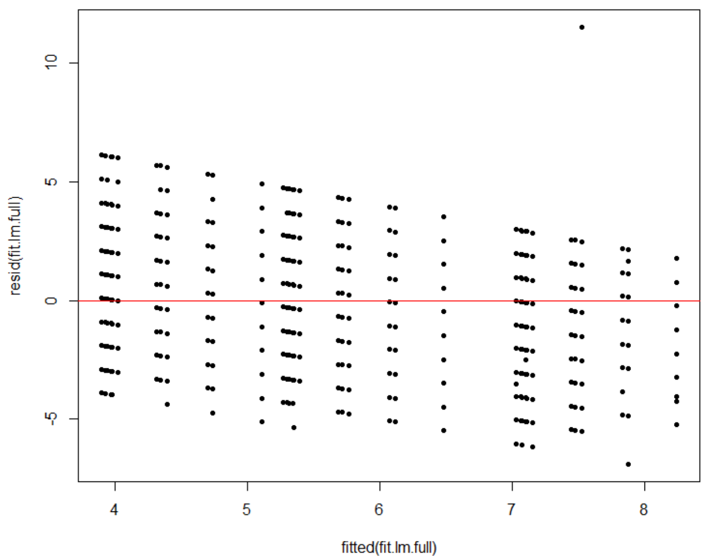
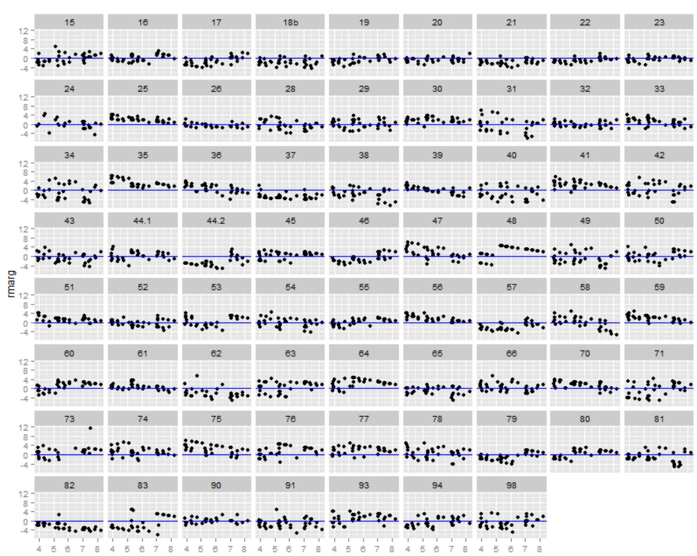
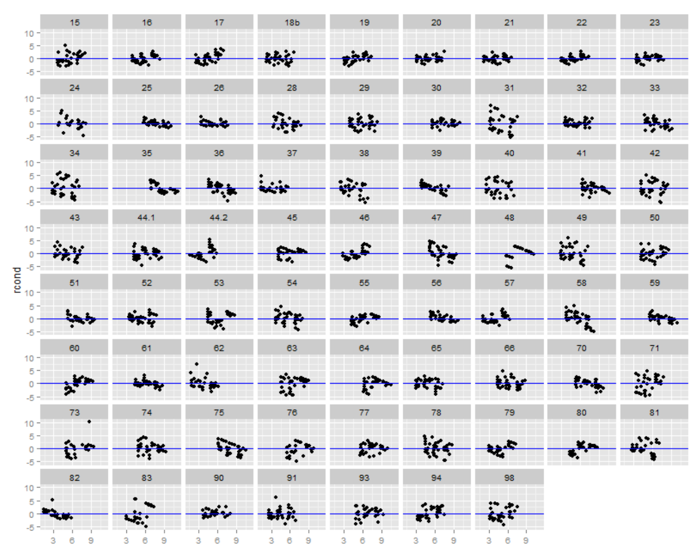


Figure: Residuals vs. fitted values for fit.lm.full





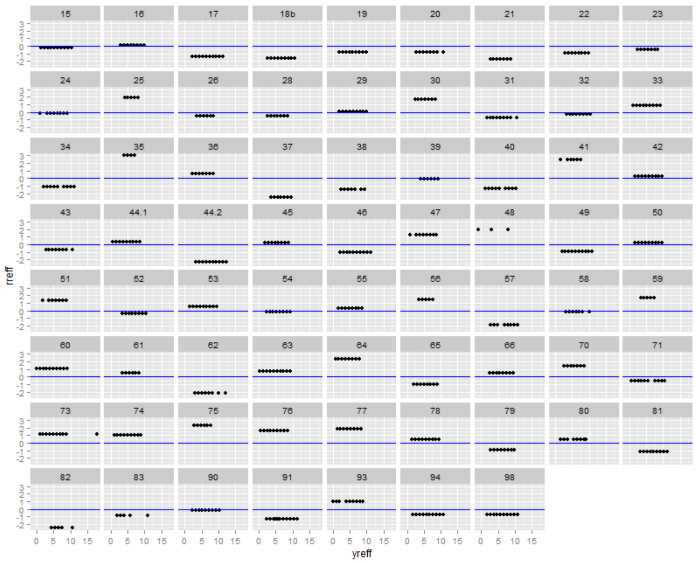


Figure: Marginal residuals vs. fitted (left, previous page), conditional residuals vs. fitted (right, previous page), and random effects residuals vs. fitted (left) for M1

In conclusion, we should include the random intercept in the model moving forward.

iii. In order to do this, we repeat the same process as in number 1 but with the random intercept term (1|subject) in the model.

```
M2 <- lmer(Classical ~ Harmony + Voice + (1|Subject))

M3 <- lmer(Classical~ Instrument + Voice + (1|Subject))

M4 <- lmer(Classical ~ Harmony+Voice + (1|Subject))

M5 <- lmer(Classical~ Instrument + (1|Subject))

M6 <- lmer(Classical~ Harmony + (1|Subject))

M7 <- lmer(Classical~ Voice + (1|Subject))
```

	M1	M2	M3	M4	M5	M6	M7
AIC	10491.51	11423.04	10552.74	11423.04	10566.14	11429.98	11461.42
BIC	10549.73	11469.6	10593.49	11469.6	10595.25	11464.91	11490.53

Both AIC and BIC prefer the first M1 model, which is the full model including the random slope for each subject. Therefore, the best model is M1.

c.

i. The new model with all three new random effect terms is the following:

```
M.n1 <- lmer(Classical ~ Instrument + Harmony + Voice + (1|Subject:Instrument) +
(1|Subject:Harmony) + (1|Subject:Voice))
```

I will compare this model (M.n1) with M1 (model from 1b) and fit.lm.full (model from 1a) using AIC and BIC measures.

	M.n1	M1	Fit.lm.full
AIC	10075.51	10491.51	11230.45
BIC	10145.37	10549.73	11282.84

Both AIC and BIC criteria prefer the new M.n1 model with three random effects varying the intercept to account for personal biases in classical ratings given categories in each of the three experimental variables. Both AIC and BIC improve by more than 300.

ii. We can repeat the same process as 1a and 1b in order to assess the influence of all three main experimental factors on classical ratings.

```
M.n2 <- lmer(Classical ~ Harmony + Voice + (1|Subject:Instrument) + (1|Subject:Harmony)
+ (1|Subject:Voice))

M.n3 <- lmer(Classical~ Instrument + Voice + (1|Subject:Instrument) +
(1|Subject:Harmony) + (1|Subject:Voice))

M.n4 <- lmer(Classical ~ Harmony+Voice + (1|Subject:Instrument) + (1|Subject:Harmony)
+ (1|Subject:Voice))

M.n5 <- lmer(Classical~ Instrument + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.n6 <- lmer(Classical~ Harmony + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.n7 <- lmer(Classical~ Voice + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))
```

	M.n1	M.n2	M.n3	M.n4	M.n5	M.n6	M.n7
AIC	10075.51	10176.17	10101.74	10176.17	10118.89	10194.3	10204.66
BIC	10145.37	10234.38	10154.13	10234.38	10159.64	10240.87	10245.41

Again, both AIC and BIC criteria prefer the first model, which is the random effects model including all three random effects for all three experimental factors.

```
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
```

The output above shows the three variance components for the random effects. We can note that the three variance estimates for each of the three experimental factors sum to the variance estimate of the total residual value. These variances quantify the personal biases in each of the three factors in the classical ratings. For harmony, the variance estimate is 0.443, this shows how much each intercept varies for an "average" subject given harmony. For voice, the variance estimate is 0.028, which shows how much each intercept varies for an "average" subject given voice. Lastly, for instrument, the variance estimate is 2.199, which shows how much each intercept varies for an "average" subject given instrument. In conclusion, this result shows that personal biases in classical ratings are the strongest for instrument and the weakest for voice.

iii. Mathematical model:

$$Classical_i = \alpha_{j[i]} + \beta_1(Harmony_i) + \beta_2(Instrument_i) + \beta_3(Voice_i) + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_j = \beta_0 + \eta_j + \gamma_j + \iota_j$$
$$\eta_j \sim N(0, \tau_1^2)$$
$$\gamma_j \sim N(0, \tau_2^2)$$
$$\iota_j \sim N(0, \tau_3^2)$$

2.

a. Before we start deciding which covariates should be included in the initial model going into the model building process, it is important to assess each variable and understand how they are coded and which of them have missing values, etc. Several of the possible predictor variables had NA values and therefore needed be taken care of in order to allow for comparison between models using selection criteria such as BIC and AIC. For some variables with missing values, it makes sense to code the NA's as zeros. For example, if someone did not answer their skill level for their 1st instrument (X1stInstr) then it is most likely that the person does not play an instrument. Therefore 0 (not at all) would be appropriate to replace NA values. Similarly, I coded in 0's for NA values for X2ndInstr, APTheory, NoClass, CollegeMusic, KnowAxis, KnowRob, Composing, ClsListen, PachListen, and ConsNotes variables. However, I am making a huge assumption by replacing NA values with 0's. In reality, I would want to consult the researcher to assess how to address the Na's, keeping in mind the three categories of missing values: Missing Completely Random, Missing at Random, Missing not at Random. I decided to exclude the variables X1990s2000s and X1990s2000s.minus.1960s1970s, since they hold missing values that I am not sure how to deal with. All other variables do not include NA's.

Furthermore, I dichotomized/collapsed some of the variables in order to make each category in the respective variables to have around equal number of observations. GuitarPlay, PianoPlay, NoClass, KnowRob, KnowAxis and Composing were turned into indicator variables (0,1). Then, I collapsed KnowRob and KnowAxis into one variable called Knowledge, which is marked 1 when either of KnowRob or KnowAxis is 1 and 0 otherwise. I collapsed categories 2+ into one level for the ClsListen variable, and therefore the new ClsListen variable has 3 levels instead of 6. The R code for making these adjustments to each variable is shown below:

```
PianoPlay2 <- ifelse(PianoPlay==0, 0, 1)

GuitarPlay2 <- ifelse(GuitarPlay==0, 0, 1)

X1stInstr2 <- X1stInstr

X1stInstr2[is.na(X1stInstr)==TRUE] <- 0

X2ndInstr2 <- X2ndInstr

X2ndInstr2[is.na(X2ndInstr)==TRUE] <-0

APTheory2 <- APTheory

APTheory2[is.na(APTheory)==TRUE] <- 0

NoClass2 <- NoClass
```

```
NoClass2[is.na(NoClass)==TRUE] <- 0 #Assume NA's are 0

NoClass2 <- ifelse(NoClass2==0,0,1)

CollegeMusic2 <- CollegeMusic

CollegeMusic2[is.na(CollegeMusic)==TRUE] <- 0

KnowAxis2 <- KnowAxis

KnowAxis2[is.na(KnowAxis)==TRUE] <- 0

KnowAxis2 <- ifelse(KnowAxis==0,0,1)

KnowRob2 <- ifelse(KnowRob==0, 0, 1)

KnowRob2 <- KnowRob2

KnowRob2[is.na(KnowRob)==TRUE] <- 0

KnowRob2 <- ifelse(KnowRob2==0,0,1)

Knowledge <- rep(0,length(KnowRob2))

Knowledge[KnowRob2==1|KnowAxis2==1] <- 1

table(Knowledge)

ClsListen2 <- ClsListen

ClsListen2[is.na(ClsListen)==TRUE] <- 0

ClsListen2[(ClsListen2>=2)==TRUE] <- 2

PachListen2 <- PachListen

PachListen2[is.na(PachListen)==TRUE] <- 0

ConsNotes2 <- ConsNotes

ConsNotes2[is.na(ConsNotes)==TRUE] <- 0

Composing2 <- ifelse(Composing==0, 0, 1)

Composing2[is.na(Composing2)==TRUE] <- 0
```

In addition, I observed that there is a skew in the variable OMSI, and conducted powerTransform() on the response and OMSI to see if x transformation would be desirable. The resulting lambda value was very close to zero, indicating that a log transformation on OMSI may be a good idea. Therefore, I will consider log(OMSI) instead of OMSI moving forward. Also, I discovered that there may be a possible interaction between log(OMSI) and Selfdeclare variables. Logically, this makes sense because those who rate themselves higher in the "are you a musician" question is more likely have a higher OMSI score, which is a test on musical knowledge. Therefore, I will include the interaction term between log(OMSI) and Selfdeclare variables in the initial model.

Therefore, I began the model selection process with the following initial model:

```
Mcov1 <- lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare + log(OMSI) +
X16.minus.17 + PachListen2 + ClsListen2+ Knowledge + CollegeMusic2 + NoClass2+
APTheory2 + Composing2 + PianoPlay2 + GuitarPlay2 + X1stInstr2 +X2ndInstr2 +
(1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice) +
log(OMSI)*Selfdeclare)
```

From this model, I took out a variable with the lowest absolute value of t while also making sure that the AIC and BIC improves at each step. I considered absolute value of t value greater than 2 as having significant effect on the response variable. The order in which I eliminated each variable is the following: PachListen2, CollegeMusic2, log(OMSI)*Selfdeclare, X2ndInstr2, APTheory2, NoClass2, Knowledge, ClsListen2, X1stInstr2, then PianoPlay2. Although log(OMSI) term turned out to be not statistically significant, I decided to keep it in because the researchers may find it important to keep it in the model, since every subject took the test and the test may have been given in order to assess their musical knowledge for this particular experiment. Then the resulting model is the following:

```
Mcov11 <-  lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare + log(OMSI) +
X16.minus.17 +  Composing2 + GuitarPlay2 +  (1|Subject:Instrument) +
(1|Subject:Harmony) + (1|Subject:Voice))
```

Every covariate predictor and the three original experimental factors are statistically significant, with the exception of log(OMSI). As mentioned earlier, I decided to keep this predictor despite the somewhat mild significance of the t value. Then, we can check AIC and BIC to see if this model actually does better than the simple repeated measures model with just three fixed effects. As shown below, AIC prefers the new model, which is noted in Mcov11 to M.n1, but the BIC prefers the simpler model M.n1. This is because BIC prefers parsimonious models and punishes the addition of predictors more severely than AIC.

	Mcov11	Mcov1	M.n1	M1	Fit.lm.full
AIC	10068.2	10092.42	10075.51	10491.51	11230.45
BIC	10167.16	10249.59	10145.37	10549.73	11282.84

I believe that having these covariates in the model may be important to the investigators, therefore I believe that the new model Mcov11 is preferable to the simpler model M.n1. However, I would need to check with the researchers to see whether or not this is a good decision.

b. In order to answer this question, we need to compare this model against the following model:

```
Mcov11.2 <- lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare + log(OMSI) +
X16.minus.17 +  Composing2 + GuitarPlay2 +  (1|Subject))
```

This model has the random intercept for each subject, but does not have the three components for the random effects for the three experimental factors.

	Mcov11	Mcov11.2
AIC	10068.2	10494.28
BIC	10167.16	10581.6

As the table displayed above shows, the Mcov11 model with three random effects intercepts does better with both AIC and BIC criteria. Therefore, we do not need to make changes in the random effect and move forward in our analysis with the model Mcov11.

c. Then the final model produces the following regression output:

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.43919  0.6627
 Subject:Voice      (Intercept) 0.02737  0.1654
 Subject:Instrument (Intercept) 1.93495  1.3910
```

```
 Residual                          2.43846  1.5616
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210;
Subject:Instrument, 210

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       4.38308    0.51256   8.551
Instrumentpiano   1.36499    0.24755   5.514
Instrumentstring  3.12835    0.24725  12.652
HarmonyI-V-IV    -0.03018    0.14280  -0.211
HarmonyI-V-VI     0.77095    0.14278   5.399
HarmonyIV-I-V     0.05612    0.14273   0.393
Voicepar3rd      -0.40721    0.08163  -4.988
Voicepar5th      -0.37110    0.08157  -4.550
Selfdeclare      -0.47913    0.13482  -3.554
log(OMSI)         0.17261    0.12023   1.436
X16.minus.17     -0.10720    0.03727  -2.877
Composing2        0.68136    0.27763   2.454
GuitarPlay2       0.84083    0.30584   2.749
```

Every covariate is statistically significant, except for log(OMSI). Both instrument factors are statistically significant and positive. Holding all other variables in the model constant, when the instrument is piano, the classical ratings are on average 1.365 higher than when the instrument is guitar. Similarly, when the instrument is string, the classical ratings are on average 3.128 higher than when the instrument is guitar. Only one level is significant in the case of harmony—harmony I-V-VI. In other words, holding all other variables in the model constant, when the stimulus has harmony I-V-VI, the classical ratings are on average 0.771 higher than when the stimulus has harmony I-IV-V. Both voice levels are significant, meaning holding all other variables in the model constant, when the stimulus has par 3rd leading voice, the classical rating is on average 0.407 lower than when contrary. Similarly, holding all other variables in the model constant, when the stimulus has par 5th leading voice, the classical ratings on average are 0.371 lower than for contrary.

Here, the selfdeclare variable is an ordered categorical variable. So with each increase in the level of the variable, there the average decrease in the classical ratings is 0.479, holding all other variables constant. Log(OMSI) has a positive coefficient, meaning the higher one's OMSI score, the higher their classical ratings are (keep in mind that this coefficient is not statistically significant). X16.minus.17 has a negative coefficient, meaning that the higher his/her score on the auxiliary test, the lower their classical ratings are. The coefficient on composing is positive, meaning that those who compose have higher classical ratings. Lastly, guitarplay also has a positive coefficient, meaning that those who play the guitar have higher classical ratings. The average intercept is 4.383, and this intercept varies by random effects shown in the first part of the regression output above. More specifically: For harmony, the variance estimate is 0.439, this shows how much each intercept varies for an "average" subject given harmony. For voice, the variance estimate is 0.027, which shows how much each intercept varies for an "average" subject given voice. Lastly, for instrument, the variance estimate is 1.935, which shows how much each intercept varies for an "average" subject given instrument. In conclusion, this result shows that personal biases in classical ratings are the strongest for instrument and the weakest for voice.

3. We can dichotomize the Selfdeclare variable using the following R code:

```
Selfdeclare2 <- ifelse(Selfdeclare<=2, 0, 1)
```

I decided that the best cutoff for dichotomizing this particular variable was 2 and above, since that would allow for a fairly even distribution of observations in the sample. The following figures display distributions of the selfdeclare variable before and after. 0 means low self-declare musician score and 1 refers to high self-declare musician score.

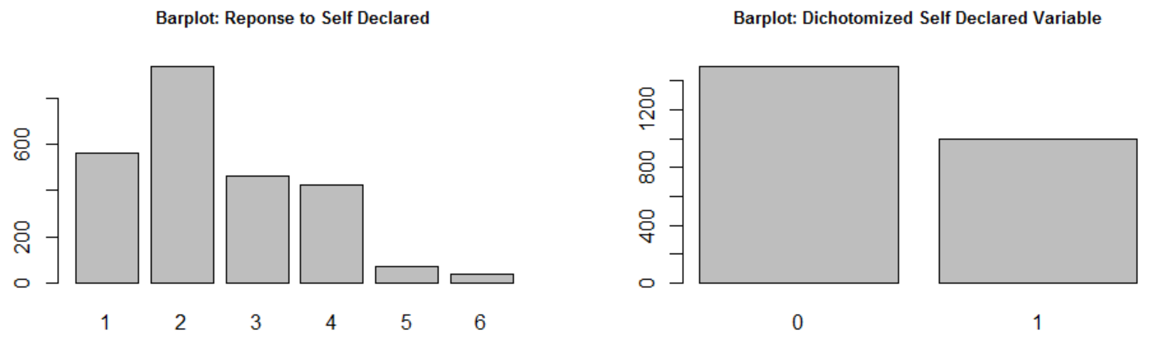


Figure: Bar plots showing the distribution of selfdeclare variable before and after dichotomization

Then I proceeded replace the selfdeclare variable in the model from 2c, which is called Mcov11. The resulting model is the following:

```
Mcov2.1 <-  lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare2 + log(OMSI) +
X16.minus.17 +  Composing2 + GuitarPlay2 +  (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))
```

From here, I included interaction terms between selfdelcare2 and every other predictor variables in the model one-by-one. From this process, I found that the interaction with x16.minus.17 is the most significant, with the t-value of -3.945.

```
Mcov2.3 <-  lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare2 + log(OMSI) +
X16.minus.17 +  Composing2 + GuitarPlay2 +  (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice) +Selfdeclare2*X16.minus.17)
```

	Mcov2.1	Mcov2.3
AIC	10077.09	10067.19
BIC	10176.05	10171.97

We see that both AIC and BIC prefers the Mcov2.3 model, which is the model including the interaction term between selfdeclare (dichotomized) and x16.minus.17 variables. The resulting regression output is the following:

```
Random effects:
 Groups              Name        Variance Std.Dev.
 Subject:Harmony     (Intercept) 0.43844  0.6621
 Subject:Voice       (Intercept) 0.02735  0.1654
 Subject:Instrument (Intercept) 1.89788  1.3776
 Residual                        2.43855  1.5616
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument,
210

Fixed effects:
                          Estimate Std. Error t value
```

```
(Intercept)                  4.02817    0.54554   7.384
Instrumentpiano              1.36591    0.24541   5.566
Instrumentstring             3.12797    0.24510  12.762
HarmonyI-V-IV               -0.03003    0.14272  -0.210
HarmonyI-V-VI                0.77097    0.14271   5.402
HarmonyIV-I-V                0.05622    0.14265   0.394
Voicepar3rd                 -0.40708    0.08163  -4.987
Voicepar5th                 -0.37118    0.08156  -4.551
Selfdeclare2                 0.13688    0.34305   0.399
log(OMSI)                    0.03372    0.11766   0.287
X16.minus.17                -0.00713    0.04554  -0.157
Composing2                   0.57509    0.27292   2.107
GuitarPlay2                  0.76024    0.30577   2.486
Selfdeclare2:X16.minus.17 -0.31392    0.07958  -3.945
```

As shown above in the regression output, the interaction term selfdeclare2*x16.minu.17 has a significant t value. Now, both the main effects are not significant, but they will stay in the model as the interaction term is significant and also improves the model AIC and BIC. The coefficient on the interaction term is negative, coefficient on selfdeclare 2 is positive, and the coefficient on the x16.minus.17 is negative. Meaning, those who declare themselves as "musicians" the greater the value of x16.minus.17, then the lower the classical ratings on average are, vice versa.

4.

a.

When only considering the random intercept (1|subject) the models are the following:

```
M.pp1 <- lmer(Popular ~ Instrument + Harmony + Voice+ (1|Subject))

M.pp2 <- lmer(Popular ~ Harmony + Voice+ (1|Subject))

M.pp3 <- lmer(Popular ~ Instrument + Voice+ (1|Subject))

M.pp4 <- lmer(Popular ~ Instrument + Harmony+ (1|Subject))

M.pp5 <- lmer(Popular ~ Instrument + (1|Subject))

M.pp6 <- lmer(Popular ~ Harmony + (1|Subject))

M.pp7 <- lmer(Popular ~ Voice + (1|Subject))
```

Then the BIC and AIC values are:

	M.pp1	M.pp2	M.p3	M.pp4	M.pp5	M.pp6	M.pp7
AIC	10453.12	11152.82	10447.49	10447.4	10441.77	11146.51	11145.87
BIC	10511.34	11199.39	10488.24	10493.97	10470.87	11181.44	11174/97

Both AIC and BIC prefer the fifth model, which is a one-predictor model with instrument. Then, we can consider all three random effects (1|subject:harmony), (1|subject:voice), and (1|subject:instrument).

```
M.p1 <- lmer(Popular ~ Instrument + Harmony + Voice+ (1|Subject:Instrument) +
(1|Subject:Harmony) + (1|Subject:Voice))

M.p2 <- lmer(Popular ~ Harmony + Voice+ (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.p3 <- lmer(Popular ~ Instrument + Voice+ (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.p4 <- lmer(Popular ~ Instrument + Harmony + (1|Subject:Instrument) + (1|Subject:Harmony)
+ (1|Subject:Voice))

M.p5 <- lmer(Popular ~ Instrument + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.p6 <- lmer(Popular ~ Harmony + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))

M.p7 <- lmer(Popular ~ Voice + (1|Subject:Instrument) + (1|Subject:Harmony) +
(1|Subject:Voice))
```

	M.p1	M.p2	M.p3	M.p4	M.p5	M.p6	M.p7
AIC	10097.24	10177.79	10089.39	10091.75	10170.06	10172.41	10170.06
BIC	10167.09	10236	10141.78	10149.96	10210.81	10218.98	10210.81

Both AIC and BIC prefer M.p3 model, which is the two predictor model with three random effects. In other words, instrument and voice are influential for popular ratings with random effects. Although in both cases non-full models perform better, for the purpose of this study, where the researchers are interested in the three experimental factors, we will be proceeding with all three fixed effects in the model. great

b. We start modeling with the following initial model:

```
Mcov3.3 <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare + log(OMSI) +
X16.minus.17 + PachListen2 + ClsListen2+ Knowledge + CollegeMusic2 + NoClass2+
APTheory2 + Composing2 + PianoPlay2 + GuitarPlay2 + X1stInstr2 +X2ndInstr2 +
(1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice) +
log(OMSI)*Selfdeclare)
```

Using the same criteria as 2a, we can eliminate covariates in the following order (one-by-one): Pianoplay2, APTheory2, ClsListen2, PachListen2, NoClass2, CollegeMusic2, X1stInstr2, GuitarPlay2, Composing2, X2ndInstr2, log(OMSI)*Selfdeclare, and x16.minus.17. What we're left with is the following:

by what criterion are you eliminating them?

```
Mcov3.4 <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare + log(OMSI)+
Knowledge +  (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice))
```

The regression output for this model is:

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41016  0.6404
 Subject:Voice      (Intercept) 0.03203  0.1790
 Subject:Instrument (Intercept) 1.89840  1.3778
 Residual                       2.49056  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210;
Subject:Instrument, 210

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       5.74243    0.49802  11.531
Instrumentpiano  -0.94814    0.24569  -3.859
Instrumentstring -2.60537    0.24539 -10.617
HarmonyI-V-IV    -0.02562    0.14046  -0.182
HarmonyI-V-VI    -0.27153    0.14044  -1.933
HarmonyIV-I-V    -0.18557    0.14038  -1.322
Voicepar3rd       0.16370    0.08320   1.968
Voicepar5th       0.16202    0.08314   1.949
Selfdeclare       0.18644    0.11763   1.585
log(OMSI)         0.05124    0.11959   0.428
Knowledge         0.49381    0.25581   1.930
```

should selfdeclare be cont or factor variable?
nice to think of log xform!

One thing to note is that I kept in selfdeclare and log(OMSI) variables since I decided that those two are somewhat important to the investigators. This is because OMSI seems to be a test of musical knowledge that the investigators were able to give to every subject. In addition, selfdeclare is a variable of interest in the next part of the analysis.

As for the interpretation of this regression output: When the stimulus has an instrument piano, the average classical rating is 0.948 lower than that where the instrument is guitar. Similarly, when the instrument is string, the average popular rating is 2.605 lower than that where the instrument is guitar, holding all other variables in the model constant. As per harmony, only I-V-VI motion has significantly different popular ratings from I-IV-V when holding other variables in the model constant. More specifically, I-V-VI motion leads to 0.272 lower popular ratings on average. Both par3rd and par5th seem to have higher classical ratings on average than contrary. Log(OMSI) has a positive coefficient, meaning that the higher one's OMSI score is, the higher their popular ratings tend to be. However, this coefficient is not statistically significant. Selfdeclare is an ordered categorical variable, so for each increase in the level of selfdeclare, there is an increase of 0.186 in the average popular ratings holding all other variables in the model constant. Knowledge is a variable that I made collapsing Know.axis and Know.rob. Because of the fact that this variable is not from the original dataset and I may be losing some information, we should be cautious when assessing the significance of this variable. However, as it stands, those who know either axis or rob have 0.494 higher average popular ratings than those do not, holding all other variables in the model constant. The average intercept is 5.742, and this intercept varies by random effects shown in the first part of the regression output above. More specifically: For harmony, the variance estimate is 0.410, this shows how much each intercept varies for an "average" subject given harmony. For voice, the variance estimate is 0.032, which shows how much each intercept varies for an "average" subject given voice. Lastly, for instrument, the variance estimate is 1.898, which shows how much each intercept varies for an "average" subject given instrument. In conclusion, this result shows that personal biases in classical ratings are the strongest for instrument and the weakest for voice.

Furthermore, we can compare this final model to the initial three predictor model with their random effects as intercepts.

	M.p1	Mcov3.4
AIC	10097.24	10100.13
BIC	10167.09	10193.27

Both AIC and BIC actually prefers the simpler model with the three predictors and the random effects. ok

c. From the model from part b, Mcov3.4, I replaced selfdeclare with the dichotomized selfdeclare2 variable, which resulted in the following model:

```
Mcov5.1 <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare2 + log(OMSI)+ Knowledge
+  (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice))
```

Then, I assessed interaction terms by adding in interaction term between Selfdeclare2 and log(OMSI) and knowledge one-by-one. I found that the t-value was significant on the coefficient of the interaction term between Knowledge and Selfdeclare2. The resulting model is:

why not try to interact selfdec2 with all fixed effects?

```
Mcov5.3 <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare2 + log(OMSI)+
Knowledge*Selfdeclare2 +  (1|Subject:Instrument) + (1|Subject:Harmony) + (1|Subject:Voice))
```

And the regression output is:

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.40966  0.6400
 Subject:Voice      (Intercept) 0.03187  0.1785
 Subject:Instrument (Intercept) 1.86712  1.3664
 Residual                       2.49074  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210;
Subject:Instrument, 210

Fixed effects:
                       Estimate Std. Error t value
(Intercept)             5.93970    0.53648  11.072
Instrumentpiano        -0.94809    0.24387  -3.888
Instrumentstring       -2.60564    0.24356 -10.698
HarmonyI-V-IV          -0.02540    0.14041  -0.181
HarmonyI-V-VI          -0.27138    0.14039  -1.933
HarmonyIV-I-V          -0.18550    0.14034  -1.322
Voicepar3rd             0.16378    0.08318   1.969
Voicepar5th             0.16209    0.08311   1.950
Selfdeclare2           -0.05257    0.31547  -0.167
log(OMSI)               0.10315    0.11602   0.889
Knowledge              -0.14405    0.37518  -0.384
Selfdeclare2:Knowledge  1.17619    0.50596   2.325
```

	Mcov5.1	Mcov5.3
AIC	10096.51	10092.68
BIC	10183.83	10185.82

After adding in the interaction term between selfdeclare2 and knowledge, the model seems to perform better by AIC means. However, BIC still prefers the model without the interaction term.

Write-up

The research question at hand is: what factors influence how classical or popular music stimuli sound to listeners? Because the data comes from a designed experiment intended to measure the influence of instrument, harmonic, motion, and voice leading on listeners' identification of music stimuli as "classical" or "popular", I will consistently consider those three experimental factors in my analyses. However, it is also important to consider other covariates and how they interact with the instrument, harmony, and voice variables.

First, I considered the influence of the three main experimental factors on classical ratings by building a repeated measures model and also considering other covariates. This model allows for personal biases within different categories of harmony, instrument, and voice through a varying-intercept by subject. As my analysis in 3 shows, the final model includes the three experimental factors (instrument, harmony, and voice), self-declared musician score (high/low), logged test score of musical knowledge (OMSI), the auxiliary score of listener's ability to distinguish between classical and popular music (x16.minus.17), whether or not the listener composes (0/1), and whether or not the listener plays the guitar (0/1). I also included an interaction term between auxiliary score and self-declared musician score. This interaction term accounts for the fact that when one has a higher auxiliary score and declares himself as a musician, they are less likely to rate the stimulus as classical. One thing to note I decided put the log(OMSI) variable in the model since I believe that this variable is somewhat important to the investigators. This is because OMSI seems to be a test of musical knowledge that the investigators were able to give to every subject.

As researchers hypothesized, instrument does have a large influence on the classical ratings. More specifically, when the instrument is string quartets compared to an electrical guitar, the classical rating goes up by 3.180 points on average, holding all other variables in the model constant. This is a large change in the classical ratings, which reinforces the researchers' belief about the influence of instruments on how "classical" stimulus sounds. Furthermore, we can conclude that stimuli with string quartets are most frequently rated as classical.

Moreover, I also agree with the researchers' claim that I-V-VI harmonic progression may be frequently rated as classical, based on my analysis in 3. The coefficient on Harmony I-V-VI is fairly large and highly statistically significant. I arrived at the conclusion that I-V-VI harmonic progression is rated as classical more frequently than any other harmonic progression since the coefficient is positive and the largest out of all harmonic progression coefficients.

Lastly, I found that the contrary motion is most frequently rated as classical out of the three voice leading categories, which also confirms researchers' belief. As noted in 3, both parallel 3rds and parallel 5ths have negative coefficients, indicating that the contrary leading voice (baseline) corresponds to higher classical ratings, while holding all other variables in the model constant.

My model for popular ratings from 4c is the repeated measures model with three experimental factors and several covariates with varying-intercept. This model reinforces my conclusions as I observed opposite relationships in the coefficients from the classical ratings model from 3. Although I used slightly different covariates (log(OMSI),selfdeclare2, knowledge, and selfdeclare*knowledge) to come up with the final model, the coefficients concerned in assessing the three hypotheses are still statistically significant. The interaction term selfdeclare2*knowledge accounts for the fact that one who has either heard of the two comedic acts or declares himself as a musician tends to give lower popular ratings than those who have heard of the two comedic acts and declare himself as a musician.

In conclusion, I was able to confirm all three of Dr. Jimenez's hypotheses about the influence of harmony, voice, and instrument on listeners' identification of music stimuli as "classical" or "popular". In addition, I was able to identify several covariates that are important to include in the analysis in order to correctly model the classical (model from 3) and popular ratings (model from 4c) given this data.

4: 18
5: 20

38

nice job