# Does Music Affect Memory?

A Controlled Experiment on a Virtual Island
Presented by the *Analysis Club* (Team No.16)

Co-authors:
*Chris Dong*
*Zhu Zhan*
*Zihao Zhou*
*Department of Statistics*
*University of California Los Angeles*

Instructor:
*Prof AM Almohalwas*
*Department of Statistics*
*University of California Los Angeles*

**Abstract**

*We have conducted a controlled experimental study with balanced and repeated measures design on the virtual Island to probe the relationship between the music and the memory. Having accounted for the effects of gender and age, our study shows that loud music, particularly the heavy metal music in this study, negatively affects the memory compared to the effect of the no treatment condition, and that the calm music, which is the classical music in this study, moderately enhances the memory compared to the effect of the no treatment condition.*

This report is prepared for the course project of Course Statistics 101B at University of California Los Angeles (UCLA) during the winter academic quarter of Year 2015. The report is last updated in 20th March, 2015. Correspondence about this report should be sent to Zihao Zhou ⟨zihao2011@g.ucla.edu⟩.

# Contents

# 1 Introduction

Memory is defined as the mental capacity or faculty of retaining and recalling certain facts as well as recognizing previous experiences. Humans have long tried to decipher the workings of our memory and to probe methods that could enhance our memories. One popular idea is that music can help one memorise. There are a lot of anecdotes or even personal experiences which appear to corroborate this belief. Indeed, most children learn the Roman alphabets by reciting them with the French melody *Ah! vous dirais-je, Maman*; some teenager students say that while they usually have a difficult time to recall the names of events and people in their history class, they often have no difficulty in managing to memorise the entire lyrics of new popular songs; Some older people also like to narrate their experience of still being able to recite the lyrics of songs popular in their childhood even though it has been decades since their childhood.

Is music just a form of entertainment, or is there actually a causal relationship between the music and the memory? Particularly, does the short-term exposure to music affects one's memory? In order to investigate this topic, we have designed and conducted a controlled experimental study on the Island, a virtual civilisation somewhere within the Pacific Ocean.

Any good study design should start with identifying the studied population and the research question. For this project, the studies population is the residents living on the virtual Island. The research question is whether music has an effect on one's memory and if so, does different music have different effect on one's memory. Before we go on to describe the structure of this project, we present a literature review into related studies.

# 2 Literature Review

In this section, we review some related studies, especially those looking into the effects of the music on the memory. To summarise existing literature on the relationship between memory and music, most studies are either about music creating an environment for memorisation attempts or acting as a mnemonic device, such as reading words in melodies. Also, there has been some academic interest about the relationship of learning or playing an instrument and memory. In this section, we introduce a study which measures the efficacy of the music as a mnemonic device.

There has been studies into studying the effectiveness of using music as a mnemonic device. Rainey and Larsen (2002) was conducted to test the hypothesis that music, in the form of familiar melody, is an effective mnemonic device. The researchers constructed two lists of 12 names which were supposed to be unknown to the subjects in the study. Participants put alternately to the two conditions, in which the names are either spoken out or sung to a familiar melody. Participants could keep listing to the spoken or musical version of the names until they are able to successfully enumerate and order the names they were given. The number of times they resumed listening to the names after they made an error were recorded. One week later, participants came back to the study and were asked to reconstruct the list of names. Again, when they made an error, they resumed listening to the spoken or sung names until they could managed to recall the list of names correctly. The researchers then recorded the number of trials participants needed to take in order to re-learn the lists. The study finds no difference between the sung and the spoken versions of the task for the initial learning but significant difference for the relearning.

There have also been hypotheses on the long-term effects of music on the memory. Some hypotheses focus not only on listening to music, but also on playing musical instruments. Ho, Cheung, and Chan (2003) tries to look at the effects of learning and playing musical instruments on both the visual and verbal memory. The study is based on the hypothesis that one can predict the impact of early experience on the development of cognitive functions if the association between the experience and the neuro-anatomy can be established (Bremner and Narayan 1998; Kolb 2014)The hypothesis is based on the findings from several studies that the life experience in life can affect both the brain structure and the cognitive functions (Hebb 1947; Wiesel, Hubel, et al. 1963). A study comparing the brain structures of people with music training with that of people without music training has found that those with former musical training tend to have a larger left planum temporale than those without musical training (Schlaug et al. 1995).

If the proposed hypothesis that changes to developments of cognitive functions can be predicted by early life experience is valid, then we should expect those with music training to have better verbal but not better visual memory, since the left temporal lobe is the primary mediator of the verbal memory (Frisk and Milner 1990; Milner 1975) and the visual memory is usually handled by the right temporal area (Saykin et al. 1992; Milner 1975). Ho et al. (2003) devises two studies to test this hypothesis. The first study is an observational study and compares and tests the differences of memory retention score between children with music training and those without any training. For the main effects of the music training treatment, the study finds statistically significant superiority in the memory retention score on verbal memory for the music training group. For the memory retention score on visual memory, the difference is not statistically significant between the two treatment groups. The second study is a follow-up longitudinal study taken one year after the first study. The second study looks into whether sustained music training will lead to greater development of memory retention. For the verbal memory, the study finds that children who continued their music training showed greater improvement in memory retention score on verbal memory than those who discontinued soon after the first study or three months after the first study.

The result of Ho et al. (2003) confirms to the neuro-anatomical hypothesis, but here we points out one problem with the design of the study: the results of the study may suffer from serious selection bias. Since it is reasonable to expect that people with better verbal memory capabilities will find it easier than those without such good verbal memory to learn to play a musical instrument, the subjects in the music group of the first study may be those who chose to learn to play the instrument because they had good verbal memory. The second study of Ho et al. (2003) has the same problem with its first study. Since it is reasonable to suspect that people with better verbal memory tend to develop their verbal memory capability faster, these people are more likely to continue their pursuit of learning to play an instrument if we assume that the better the verbal memory, the easier it is to learn to play the instrument. Of course, it would be unwise to dismiss the results of Ho et al. (2003), since the problem we have identified is indeed extremely difficult to address for every similar study. Moreover, the study has shown a vigorous attempt to isolate the causal effect of learning music on working verbal memory with its second follow-up longitudinal study.

# 3 Introduction to the Island Project

Our study differs from the two previous studies described in the literature review because in our study music serves as an environment, not as a mnemonic device as in Rainey et al. (2002) nor as a long-term life experience as in Ho et al. (2003). Moreover, we are not only interested in whether the music has an effect on the memory, but also we would like to know whether different music types can have different effects on the memory if such effects exist. With these objectives, we introduce a controlled experiment on the virtual Island.[1]

## 3.1 Factors in this Study

**Response Variable**

For the purpose of this study, the response variable should be a variable that indicates how good one is at remembering things. In previous formal academic study, memory game or test is often used to measure one's ability to memorise. Fortunately, the Island system does offer three types of memory tests, which are vocabulary test, memory game and memory cards. For the vocabulary test and memory card test, the result of the test is the proportion of correctly recalled items; the memory game is the number of seconds a player needs to match 30 pairs of cards. We have chosen the performance in the memory game as the response variable because it is a continuous variable which is suitable for the later analysis of variance. Moreover, we believe that the memory game is better at differentiating people with different memorising capability. For example, it is likely that two men with different level of memories have a perfect score on memory card test because at best they could recall 10 cards, the total number of cards in a memory card tests. On the other hand, in the memory game, we can expect someone with a better memory to use fewer seconds to complete the game.

Having defined the response variable, we should also discuss when to take measurements of it. Clearly, for this study, we should take the measurement of a participant immediately after he has been given whatever treatment.

**Treatment Factor**

There is one treatment factor in this project, the music type. We wish to have three levels of music type, no music, the calm music and the loud music. For this study, we use the classical music to represent the calm music and the heavy metal music to represent the loud music. On the Island system, each set of music involves having the subject listening to a certain kind of music for 10 minutes. For example, if a participant receives two sets of classical music, then it means that the subject listens to the classical music for 20 minutes.

**Nuisance Factors**

We have considered several nuisance factors which could have an important impact on one's memory. Such factors include one's medical conditions, the time of the day when the treatment is administered and how long is the music a participant is exposed to. As we are not directly interested in the effects of these factors, we hold all the aforementioned nuisance

[1]The Island is a virtual system designed by the Heron Island Research Station by the University of Queensland Australia and for students learning experimental design.

factors constant in our study. For the time of the day nuisance factor specifically, We will assign one treatment and the corresponding memory game at around 14 o'clock PST.

Another nuisance factor is the order of the treatment if we are going to have a repeated measures design. On one hand, we would like the order of the treatment to be determined randomly; on the other hand, we would like the overall orders of the treatments to be balanced. The method to achieve this balance while ensuring randomisation will be described in Section 5.

Then there are the demographic factors such as age and gender. Since we are not aware of serious academic evidence supporting the claim that gender plays a role in one's memory, we do not expect gender to have a significant main effect. However, we are still including gender in our study because we believe that the familiarity and attitude towards different types of music will have an impact on one's performance while one is exposed to certain type of music. Unfortunately, there is no way we can measure this cultural factor. Fortunately, we have strong reason to believe that the cultural factor is strongly associated with gender: indeed, we often expect more men than women to listen to heavy metal music. Therefore, in the study, although we do not expect the main effects of gender to be statistically significant, we still think it possible that as the substitute for the unmeasurable cultural factor, its interaction effects with the treatment factor may be significant.

For age, we have decided to have three age groups, the young, the medium and the old. The young consists of those between 11 and 30 years old; the medium consists of those between 31 and 50 years old; the old consists of those between 51 and 70 years old.

## 3.2 Design Structure

We have decided to adopt a split plot/repeated measures design because we suspect that performance between different individuals vary a lot. Moreover, due to time constraint for this project, having a repeated measures design is more efficient as it reduces the number of sample size needed, which we will discuss in the following subsection.

There are two sizes of units in our design. The larger size is each subject, which also acts as his block in this study. There are two between-block factors in this study: the age group and the gender. We also include their interaction term into our model. Within each block, or participant, the smaller size unit is each order of treatment of each participant, and the within-block factor in the music type factor. To summarise the structure of our study, we include the factor digram, with the degree of freedom of each factor, in Table 3.1 on Page 6. In the factor digram, $n$ is the number of residents in our sample.

From the factor diagram, we see that for the design to work we need to have $2(n-4) > 0$. Moreover, if we are to have a balanced design, then the sample size must be the multiple of the number of levels of each between-block factor. Therefore the minimum sample size for our design is 12.

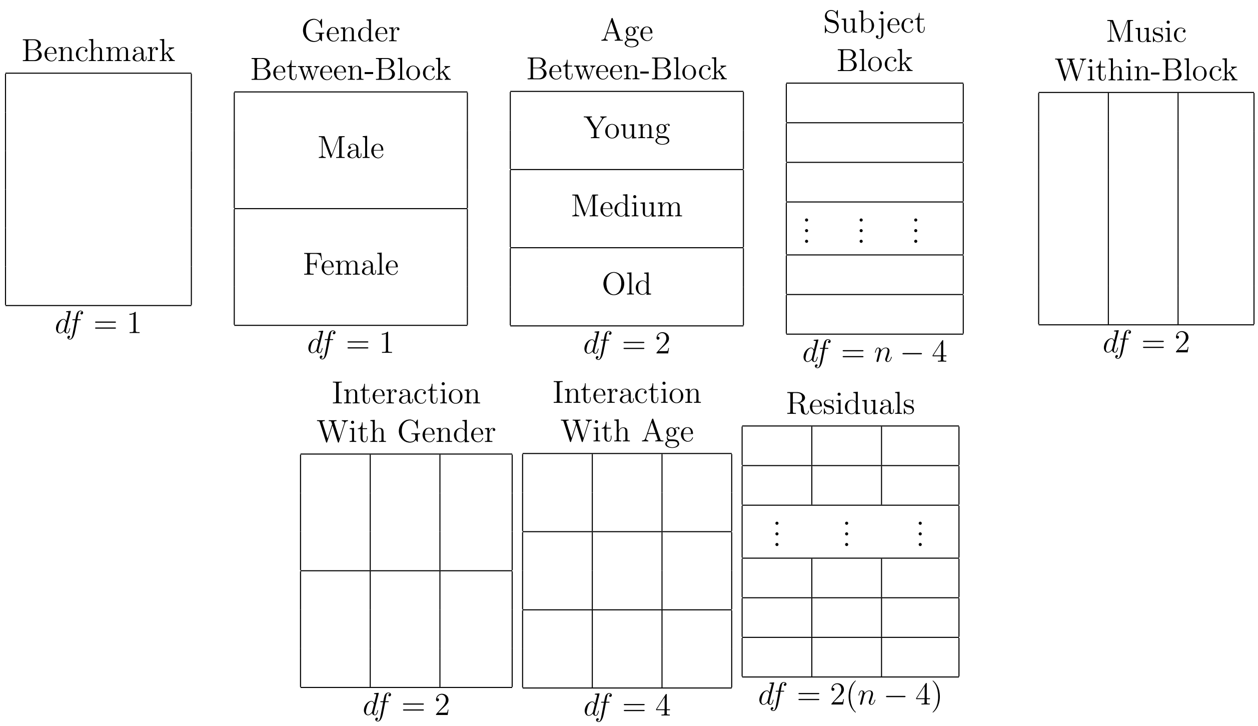



Table 3.1: Factor Diagram for the Design of the Project

# 4 Power Analysis

We run a power analysis to determine the appropriate sample size. We are most interested in the results of the pairwise comparing each pair of levels of the treatment factor. Because of the repeated measures design in our project, the appropriate hypothesis test for the treatment factor is paired t-test. We assume effect size to be 0.45 standard deviations of the test statistic, the significance level to be 80% and the alternative hypothesis is two-sided. The appropriate sample size is given in the following R output.

```
     Paired t test power calculation

              n = 40.7223
              d = 0.45
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*
```

Please note that each subject is paired with himself. Therefore the minimum required sample size for our desired power is 41. Since we strive for a balanced design, our final sample size is settled at 42, with 7 women and 7 men in each of the three age groups.

# 5 Methods

In this section, we focus on the implementation of the study, which could be roughly divided into sampling, administering treatment and collecting data.

## 5.1 Sampling

We collected our sample from the residents on the virtual Island. We do have the necessary demographic information to generate a random sample: as the system shows the population of each area, we may use a two-stage cluster sampling technique: the Island is partitioned into villages, we first randomly choose a village whose respective probability is proportional to the population of that village. In the second stage, we randomly choose one subject from that village, with each subject having equal village. If a resident does not meet the requirement for that subject, we terminate this step and repeat the procedure. For example, if age of the resident is in none of the age groups or he is already in the sample, we terminate this step and repeat the step again. We repeat this step until we have completed the goal of our sampling: 7 men and 7 women from each of the three age groups.

In practice, however, this technique is time consuming and assumes that there are no demographic changes such as migrations and deaths during the sampling period. Therefore in the end we opt for a convenience sample, which should be enough for the purpose of study, which is to probe the relationship between music and memory.

Each group member is responsible for sampling one particular age group (young, medium or old), which contains 14 participants (7 males and 7 females). Please note that before the experiment, we decide to choose subjects with no major illnesses to keep illness as a held-constant variable. However, we soon realised that it is not easy to find old people without medical conditions, since people become more vulnerable to diseases when they grow old. As a result, our final sample has several participants in the old group with medical conditions such as diabetes and asthma.

Another difficulty we encountered while sampling was that not all Island residents we got in touch with agreed to take part in our study. We did not record the exact acceptance rate of all the residents we asked for participation, but for the purpose of this project, we assume that the self-selection bias is too small to greatly impact the results of this study.

## 5.2 The ID System in This Project

We consider it worth mentioning the ID system we have used in this project. Each participant in our project is assigned with an 8-digit ID. We partition the ID and the meaning of each partition in Table 5.2 on Page 8. For collecting the data, the ID helps us check whether the information recorded in that row is correct; for analysing the study results, the ID acts as an indicator variable for the participant.

## 5.3 Treatments and Measurements

As aforementioned in the description of our study, each participant in our study will receive all levels of the treatment factor. Each participant is expected to receive one treatment at

| Digit Position | Explanation |
|---|---|
| 1 | Identifier of the collector |
| 2 | Gender. 0 = Female, 1 = Male |
| 3 | Age group. 1 = young, 2 = medium, 3 = old |
| $4-5$ | Index in the condition group, from 1 to 7 |
| $6-8$ | Treatment order. 1 = control, 2 = calm, 3 = loud |

Table 5.2: Explanation of the ID System

a time and rest for at least 24 hours before the next treatment is given. The order of the treatment is indicated as the last three digits of this participant's assigned ID. For example, if Nathan McCarthy has ID 21205123, then the order of the treatments for McCarthy is control, calm and loud.

As we are aware that the order of treatment could be a nuisance factor since the study involves playing the same memory game for three times, we randomised the order of the treatments for each participant, while keeping the overall order of each treatment balanced. We randomly assigned the 42 rows of fourteen $3 \times 3$ Latin square to the 42 participants so that for each subject, the treatment order is randomly determined but for the entire sample, each treatment is given in each order for exactly 14 times.

For the control treatment, we directly measure and record the participant's performance in the memory game; for the calm music treatment, we give two sets of classical music, each set lasting 10 minutes, to the participant and then asked the subjects to do the memory game and recorded his performance in seconds immediately after he finished listening to the music; for the loud music treatment, we used heavy metal music, with the procedure similar to the calm music treatment. By our study design, we tried to give each treatment for each participant at the same time of the day so that we can eliminate the effects of the time of the day on memory.

For the nuisance factors such as age, gender and medical conditions, we also collected each participant's date of birth, age, gender and medical locations. To help the interested reader replicate or examine our study, we have also recorded the name of the village where each participant lives.

The final dataset we have collected for the study is in wide-format and consists of 42 participants with their ID, gender, date of birth, age, age group, home location, medical conditions and the three levels of the treatments. The dataset was then converted to long format for the purpose of quantitative analysis which will be described in detail in the following section.

# 6 Results and Discussions

We are now ready to analyse the data collected and probe the research question of this project. The analysis will commence first with straightforward graphical analysis of the collected data. Then the analysis will proceed to quantitatively answer the question whether music has an effect on one's memory through repeated measure analysis of variance (rAnova).

## 6.1 Graphical Analysis

**Overall Distribution of the Response**

We now take a look at the differences between different levels of the between-block factors in this project. We first take a look at the distribution of the response variable over all sub-units in this study, which are composed of each treatment administered to each subject. The conventional five-number summary statistics of the response variable, the number of seconds needed to complete a memory game, is reported in Table 6.3 on Page 9. From the summary statistics, we see that the response variable in our study is slightly skewed to the right. What is worth noting is the range of the response variable in this project: the worst performance is more than 100 seconds longer than the best performance. We further confirm the skewness and the range of the response variable in the histogram with its density curve reported in Figure 6.1 on Page 9. The histogram shows that a majority of the response values lie between 40 to 60 seconds, conforming to our own impressions while collecting the data.

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 29.90 | 48.55 | 59.05 | 62.93 | 75.30 | 132.00 |

Table 6.3: Summary Statistics of the Response Variable

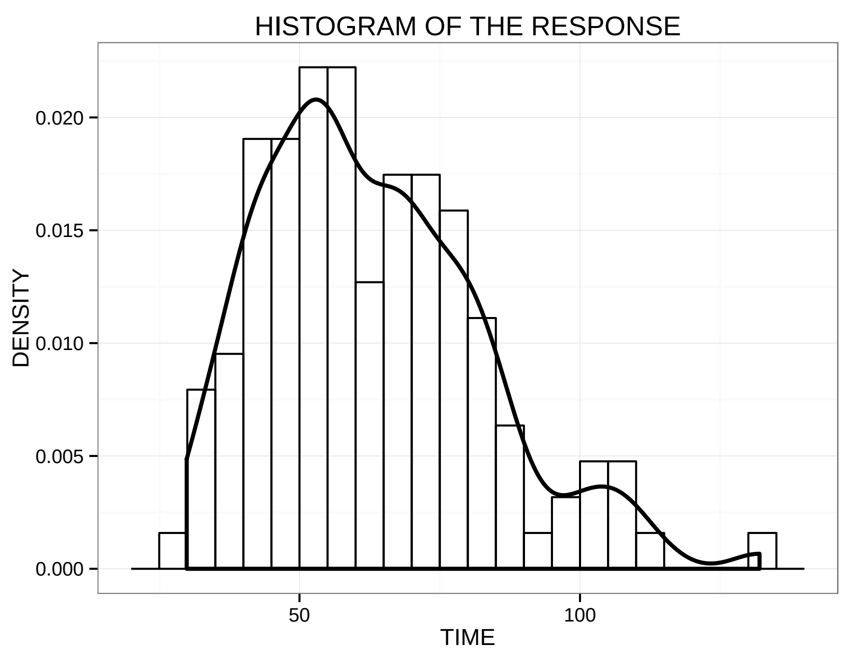



Figure 6.1: Histogram of the Number of Seconds Needed to Complete a Memory Game

**Differences Over Gender**

We first take a look at the difference between the two genders. Although as far as we are concerned, there has not rigorous academic evidence suggesting that gender plays a role in a person's ability to memorise, this factor may still be useful in this project in that different genders may behave differently under different music type due to cultural factors. The box plot of the response variable grouped by the two genders is reported in Figure 6.2 on Page 10. From the box plot, we see that the median value of the response for males is slightly larger than that for females. The difference, however, is not strong enough to determine whether this is a statistically significant difference without further quantitative analysis.

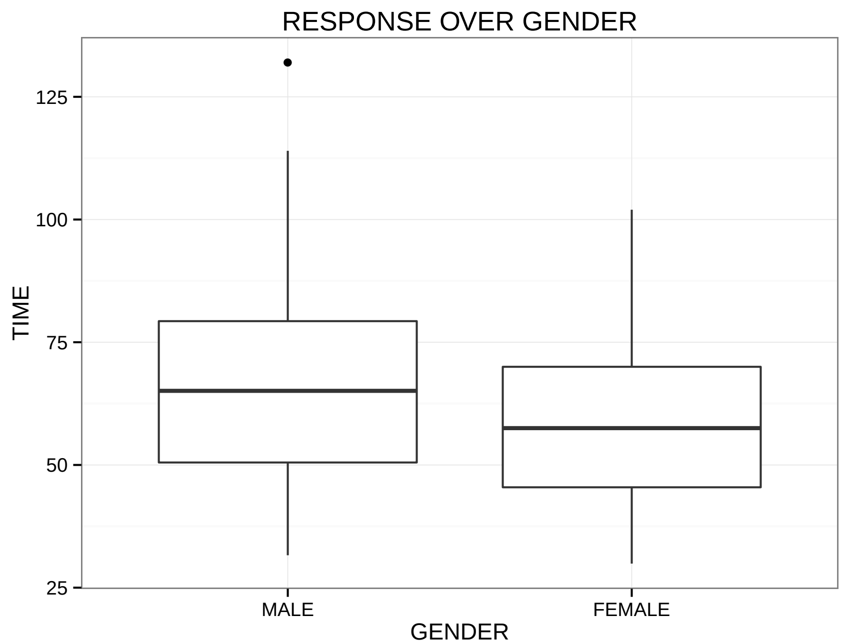



Figure 6.2: Distribution of the Response Grouped by Gender

**Differences Over Age Group**
Now we look at the distribution of the response over different age groups. As aforementioned, there are three age groups in this project. It has been a popular anecdotal belief that younger people tend to have a better memory than are older people, with which the three members in the project group also agree, but we are not sure whether there has been formal academic evidence in support of this belief. We report the box plot of the response variable in Figure 6.3 on Page 11. To our surprise, overall older people appear to have a better performance in memory game than do younger people, although we are not sure yet whether this difference is statistically significant.

**Difference Over Treatments** We now proceed to the graphical analysis for the main treatment factor, the music type. Although we are not aware of any academic study which tries to answer this project's research question, anecdotal evidence and personal experience make the authors' believe before any analysis that the louder the music, the worse the performance in the memory game tends to become. To see whether this inclined belief is

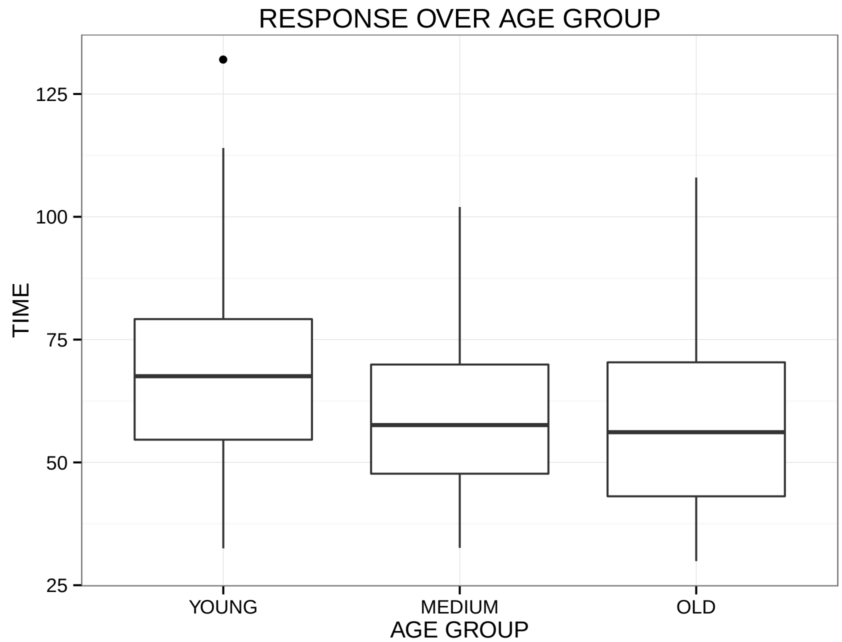



Figure 6.3: Box Plot of the Resonse Grouped by Age

supported in this project, we make the box plot of the response variable over the three treatments in this project, which are no music (control), classical music (calm) and heavy metal music (loud) in Figure 6.4 on Page 12. The figure does show a clear deterioration in the performance of the memory game for the loud music group. Also, the calm music group tends to have the best performances amongst three treatment groups, seeming to confirm the popular believe that soothing music, especially classical music, improves one's memory.[2] However, although visually detectable, the difference between the calm music group and the control group is not clear enough to deem it statistically significant without further quantitative testing.

**Interactions Effects**

For the purpose of answering the main research question, in this project, we are in fact more interested in the interaction effects of each of the two between-block factors and the treatment factor than in the main effects of these between-block factors. We expect that for the same music type, particularly the heavy metal music, older people should suffer greater deterioration of performance in the memory game. For the gender factor, we expect that women should have a more prominent deterioration in the memory game than men when exposed to loud music, since popular impression associated with the heavy metal music is that more men than women, and more youngsters than the elderly, listen to heavy metal music and thus these young men are more used to the loud music.

[2]More specifically, many people seem to believe that music composed by Beethoven, Mozart and Debussy is the most effective in enhancing one's memory. As on the Island virtual world, we cannot choose a specific composer's work for the subject to listen, we have decided to assume that all classical music has a uniform effect on one's memory, if there is any.

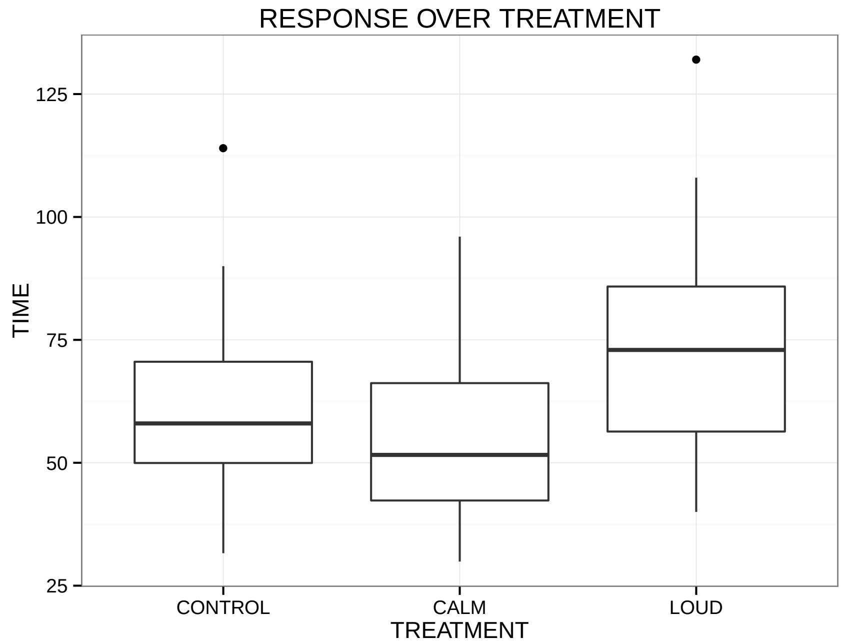
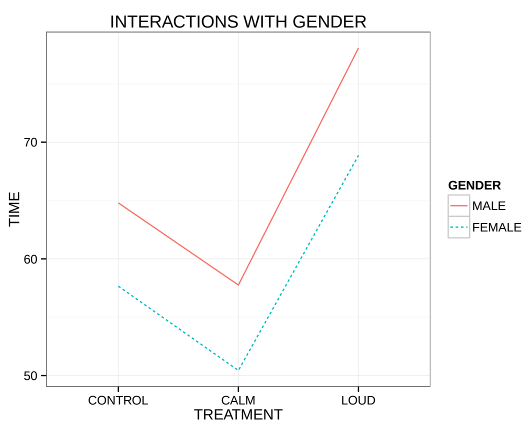
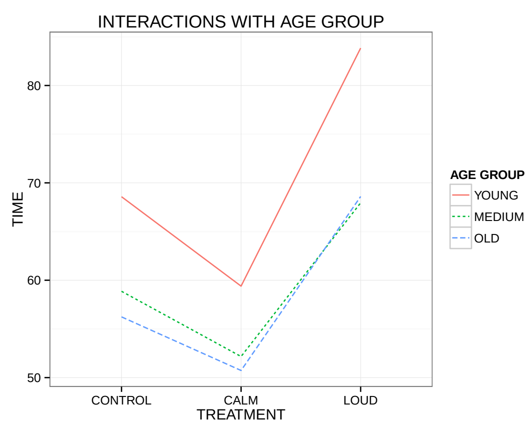



Figure 6.4: Distribution of Response Over Treatment Groups

The two interaction plots, of the interactions of the treatment with the two between-block factors, are reported together in Figure 6.5 on Page 12. From the plot, we see that the interaction effects overall are very weak. The only visually detectable interaction effect comes from the different slope between the medium and old age groups when it moves from the calm music to the loud music. This difference, however, is not clear enough for us to be confident of its statistical confidence before further analysis.




(a) Interactions with Gender




(b) Interactions with Age

Figure 6.5: Interaction Plots with Between-Block Factors

## 6.2 Quantitative Analysis

In this part we use rAnova and hypothesis testing to quantify the effects of each level of the factors in this project and the respective statistical significance. First, we specify the model in our project.

**Model Specification**
We use $i = 1, 2, 3$ to index the three levels of the treatment factor in this project, $j = 1, 2$ for the two genders, $k = 1, 2, 3$ for the three age groups, $l = 1, 2, \dots, 42$ for the 42 different Island residents participating in our project. The formula for our model is given as

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + \gamma_{jk} + \delta_l + \zeta_{ij} + \eta_{ik} + \epsilon_{ijkl}, \tag{6.1}$$

where $y_i$ is the response variable, the number of seconds used to finish the memory game; $\alpha_i$ the treatment effect of treatment $i$; $\beta_j$ the effect of female for $j = 0$ and the effect of male for $j = 1$; $\delta_k$ the effect of k'th age group; $\gamma_{jk}$ the interaction effect of gender $j$ and age group $k$; $\delta_l$ the block effect or the effects coming from the participant himself; $\zeta_{ij}$ the interaction effect of treatment $i$ and gender $j$; $\eta_{ik}$ the interaction effect of treatment $i$ and age group $k$; lastly, we have $\epsilon_{ijkl}$ as the error term for observation whose index is $ijkl$.

Assumptions for our model include that $\epsilon_{ijkl} \sim N(0, \sigma)$ for all observations and these observations are independent from each other for different $l$. Within each $l$, we do not assume independence since we have taken repeated measures on same subjects. Also, the sum of the effects of all levels of any factor in the model should add up to 0, ie $\sum \alpha_i = 0$.

**Correction for Multiple Hypothesis Testing**
As we are going to make multiple hypothesis testing, it is essential that we should adjust the individual significance level for each test in order to control the family-wise Type I Error. As the test statistics in rAnova and the pairwise comparisons are not statistically independent, we only consider the individual significance level for the pairwise comparisons.

First suppose that all factors in the model are significant and we wish to compare every pair of levels of every factor in the project. For the gender factor, there will be one confidence intervals; for the age group factor, 2 confidence intervals; for the treatment factor, there are 2 confidence intervals. Suppose the family-wise Type I Error is 0.05, using Bonferroni correction, each individual significance level should be $0.05/5 = 0.01$ and therefore the confidence level for each confidence interval should be 99%.

Please note that we have skipped the confidence interval for the interaction effects because the previous graphical analysis has shown that even in the case where the interaction effects are statistically significant, their magnitude is very small since the slope between different interaction terms changes very little.

**Repeated Measures Analysis of Variance**
In order to test for the statistical significances of each factors, we run an rAnova whose model is based on Formula (6.1). In order to test the significance of the between-block factors, each subject will act as an error term; with each subject, each treatment will then act as the error term to test the within-block treatment factor and its interactions with the between-block factors.

We run the rAnova with two error strata in R and get the summary output of the rAnova as below.

```
Error: ID
               Df Sum Sq Mean Sq F value Pr(>F)
MALE            1   1959  1958.8   2.278  0.140
AGE_GROUP       2   3739  1869.7   2.174  0.128
MALE:AGE_GROUP  2   2672  1335.8   1.553  0.225
Residuals      36  30956   859.9

Error: ID:TREATMENT
                    Df Sum Sq Mean Sq F value  Pr(>F)
TREATMENT            2   8061    4030 194.520 < 2e-16 ***
MALE:TREATMENT       2     27      14   0.656 0.52177
TREATMENT:AGE_GROUP  4    313      78   3.776 0.00745 **
Residuals           76   1575      21
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the summary above, we see that in the first error stratum where each subject acts as an error term, we see none of the two between-block factors and their interaction term have statistically significant effects. Overall this result is consistent with our previous analysis.

We then look at the second error stratum where each treatment within each subject is the error term. From the rAnova table, we see that the treatment factor, as well as its interaction with the age group, shows statistically significant effects. The two p-values associated with the two factors are also small enough to be rejected at our pre-determined significance level of 0.01.

As the treatment factor has more than 2 levels, the above rAnova assume sphericity of the treatment variable. We here run a Mauchly's sphericity test (Mauchly 1940) to see whether the sphericity assumption is violated in our dataset.

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                SS num Df Error SS den Df      F    Pr(>F)
(Intercept) 498961      1    39325     41 520.21 < 2.2e-16 ***
TREATMENT     8061      2     1915     82 172.60 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

          Test statistic p-value
TREATMENT        0.89673 0.11305
```

```
Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

          GG eps Pr(>F[GG])
TREATMENT 0.9064  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

             HF eps   Pr(>F[HF])
TREATMENT 0.9459417 1.268884e-28
```

For our purpose, please pay attention to the part of the Mauchly Test. The null hypothesis of this test is that the sphericity assumption is respected. Given that the p-value associated with the test is more than 0.1, we fail to reject the null hypothesis and therefore accept the statistical significance of the treatment factor in the previous rAnova table as valid.

Having determined whether the effect of each factor is statistically significant, we proceed to see the effect of each level of each factor in the model. Using R to summarise the effects table of the model, we report the table below.

```
Tables of effects

 MALE
MALE
 FALSE   TRUE
-3.943  3.943

 TREATMENT
TREATMENT
CONTROL    CALM    LOUD
 -1.702  -8.833  10.536

 AGE_GROUP
AGE_GROUP
 YOUNG MEDIUM    OLD
 7.676 -3.269 -4.407

 MALE:TREATMENT
       TREATMENT
MALE    CONTROL CALM    LOUD
  FALSE  0.3738  0.2810 -0.6548
  TRUE  -0.3738 -0.2810  0.6548

 TREATMENT:AGE_GROUP
         AGE_GROUP
```

```
TREATMENT YOUNG   MEDIUM  OLD
  CONTROL -0.3238  0.9071 -0.5833
  CALM    -2.3786  1.3381  1.0405
  LOUD     2.7024 -2.2452 -0.4571

 MALE:AGE_GROUP
       AGE_GROUP
MALE    YOUNG  MEDIUM OLD
  FALSE -3.290 -3.221  6.512
  TRUE   3.290  3.221 -6.512
```

The effects table can be considered as the quantitative version of the graphical analysis we have done. Having determined whether each factor is statistically significant, we jump straightly to the two factors whose effects are statistically significant.

We first look at the effects table for the treatment factor. The table shows that on average, the loud group spend 12.238 more seconds than the control group and 19.369 more seconds than the calm music group to complete the memory game. For the control and the calm music group, the control group spend 7.131 more seconds than the calm music group to finish the memory game. While we are fairly confident that the difference between the calm music and the loud music group is significant, the other two differences, particularly the differences between the control and the calm music group, require further analysis to tell whether the differences are statistically significant.

Then we look at the effects of the interactions of the treatment factor and the age group factor. From its effects table, we can tell that the differences between different levels of this factor are very small: the maximum difference in the table is about 5 seconds, which are very small compared to the extent of fluctuations in the response variable, whose overall standard deviation is around 20. Given that the estimated differences between different levels of the interactions of the treatment and the age, we do not think that it is necessary to compute the confidence interval for each pairwise difference of the levels of this interaction factor.

Therefore we only need to compute the confidence intervals of the differences between control and calm groups, between control and loud groups, and between calm and loud groups. As each participant in our project has received all three treatments whose order for the subject has been pre-determined, in order to test the difference between any two levels of the treatment factor, we need to use paired t-test. As we will correct the individual confidence level to compensate for increased Type I Error in multiple hypothesis testing, the confidence level of each confidence interval is set to 99%.

We first look at the confidence interval for the difference between the control and the calm group. As mentioned, we test the difference between the 42 observations of the response under the control condition and those under the calm condition with paired t-test.

```
Paired t-test

data:  island["CONTROL", TIME] and island["CALM", TIME]
t = 7.9796, df = 41, p-value = 7.078e-10
alternative hypothesis: true difference in means is not equal to 0
```

```
99 percent confidence interval:
 4.717057 9.544848
sample estimates:
mean of the differences
               7.130952
```

From the t-test result, we see that the the calm group performs statistically significantly better than the control group. The estimated difference is about 7.13 seconds; the lower bound at 99% confidence interval for this difference is 4.72 seconds; the upper bound is 9.54 seconds.

Next we look at the confidence interval between the control and the loud group. The output from R is printed below.

```
Paired t-test

data:  island["CONTROL", TIME] and island["LOUD", TIME]
t = -11.564, df = 41, p-value = 1.745e-14
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -15.096741  -9.379449
sample estimates:
mean of the differences
               -12.2381
```

From the output, we see that the control group performs significantly better than the loud group. The estimated superiority is about 12.24 seconds; the lower bound of this difference at 99% confidence level is about 9.38 seconds; the upper bound is about 15.10 seconds. It is clear that the difference between the control and the loud is greater than between the control and the calm group.

Lastly, we look at the difference between the calm and the loud group, with the output printed below.

```
Paired t-test

data:  island["CALM", TIME] and island["LOUD", TIME]
t = -16.2692, df = 41, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -22.58489 -16.15320
sample estimates:
mean of the differences
              -19.36905
```

The output shows that the the calm group performs significantly better than the loud group. The estimated difference is about 19.37 seconds; the upper bound of the difference at 99% level is 22.59 seconds and the lower bound is 16.15 seconds. Clearly, the difference between the calm and the loud group is the biggest amongst the all three confidence intervals.

# 7 Conclusions

Our controlled experimental study on the virtual Island has indicated that music does have a statistically significant effects on one's memory. In particular, the results form our fixed-effects model has implies that the calm music, which is represented by classical music in our project, gives the best improvements in one's memory, and that the loud music, represented in this project as the heavy metal music, affects one's memory most negatively. As we are aware of other nuisance factors that could affect one's memory, we have also introduced some nuisance factors into our model. For example, we have blocked the subjects by gender and by three pre-determined age groups. Our analysis has shown that none of the main effects of these nuisance factors are statistically significant, but the interactions effects of the treatment and the age group are statistically significant, although our graphical analysis as well as the later quantitative analysis has demonstrated that magnitude of the effects is very small.

It should be pointed out and emphasised here that we do not consider the design of our study as flawless. Indeed, given more time and resources, we may have had an even better study design. To start with, although we have tried to hold the health condition of each subject constant, not all subjects in the old age group are healthy because as we have explained, people become more likely to be sick when they get older. Moreover, although thanks to the controlled condition and the balanced design, our study does help illustrate the relationship between music and the memory, the results may not hold for the general population, because our sample was a convenience sample and therefore not representative of the entire Island population.

The main focus of our study is the effect of music type on memory, but we understand the amount of exposure to the music may also play a part. Due to the constraint of the time, our study does not study the effect of the amount of exposure. Therefore future studies about the relationship between music and memory could consider introducing the level of exposure to the study design.

# 8 Acknowledgements

We would like to thank the Department of Statistics at UCLA for offering this interesting and important course. Our gratitude also goes to Prof AM Almohalwas and Ms Jiaying Gu at the UCLA Department of Statistics, both of whom have been kind to offer their time and have given us useful advice for this project. We would like to specially thank the *Heron Island Research Station* at the *University of Queensland, Australia* for the fascinating virtual Island system and we here offer our best wishes and great anticipations of future developments on this virtual island. Lastly, we would like to thank each team member in this project for committing great amount of time and efforts to designing, conducting and analysing this project. The project would not have been this well but for the efforts and coordination of the team members.

# References

Bremner, J Douglas and Meena Narayan (1998). "The effects of stress on memory and the hippocampus throughout the life cycle: Implications for childhood development and aging". In: *Development and psychopathology* 10.04, pp. 871–885.

Frisk, Virginia and Brenda Milner (1990). "The role of the left hippocampal region in the acquisition and retention of story content". In: *Neuropsychologia* 28.4, pp. 349–359.

Hebb, Donald O (1947). "The effects of early experience on problem solving at maturity". In: *Am Psychol* 2, pp. 306–307.

Ho, Yim-Chi, Mei-Chun Cheung, and Agnes S Chan (2003). "Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children." In: *Neuropsychology* 17.3, p. 439.

Kolb, Bryan (2014). "Brain plasticity and behavioral change". In: *Advances in psychological science* 2, pp. 115–144.

Mauchly, John W (1940). "Significance test for sphericity of a normal n-variate distribution". In: *The Annals of Mathematical Statistics* 11.2, pp. 204–209.

Milner, B (1975). "Psychological aspects of focal epilepsy and its neurosurgical management." In: *Advances in neurology* 8, p. 299.

Rainey, David W and Janet D Larsen (2002). "The effect of familiar melodies on initial learning and long-term memory for unconnected text". In: *Music Perception* 20.2, pp. 173–186.

Saykin, Andrew J et al. (1992). "Neuropsychological changes after anterior temporal lobectomy". In: *The neuropsychology of epilepsy.* Springer, pp. 263–290.

Schlaug, Gottfried et al. (1995). "Increased corpus callosum size in musicians". In: *Neuropsychologia* 33.8, pp. 1047–1055.

Wiesel, Torsten N, David H Hubel, et al. (1963). "Single-cell responses in striate cortex of kittens deprived of vision in one eye". In: *J Neurophysiol* 26.6, pp. 1003–1017.

# A R Codes for This Project

We print the R codes used in this project. For the codes to run without problems, the reader should have the comma-separated-value file which is uploaded to CCLE ready in the working directory.

```
# codes for the Island project
# Please note that this script requires the uploaded .csv file to work properly

# ---------------- script_setting --------------
# library declarations
library("stringr")
library("gdata")
library("reshape2")
library("plyr")
library("data.table")
library("ggplot2")
library("lubridate")
library("xtable")
library("car")
library("pwr")

# ggplot2 theme
theme_set(theme_bw())

# ---------------- treatment_order -------------
# set seed
set.seed(7777777)
treat_order <- rep(c(123, 231, 312), 14)
# randomised order
ran_order <- sample(treat_order, size = 42, replace = F)
# order for each one
chris <- ran_order[1:14]
zoe <- ran_order[15:28]
harry <- ran_order[29:42]
# id for mine
harry_id <- str_c("3", rep(c(0, 1), c(7, 7)), 3, 0, 1:7, harry)

# -------- power_analysis --------------
# power analysis for treatment
power_analysis <- pwr.t.test(d = 0.45, sig.level = 0.05, power = 0.8,
                             type = 'paired', alternative = 'two.sided')

# -------- data_import -----------------
# Import data
```

```
island <- read.csv("island_team_16.csv", header = T, stringsAsFactors = F,
                   colClasses = c("ID" = "factor", "AGE_GROUP" = "factor",
                                  "TREATMENT" = "factor", "MALE" = "factor",
                                  "DOB" = "Date"))
island$AGE_GROUP <- factor(island$AGE_GROUP,
                           levels = c("YOUNG", "MEDIUM", "OLD"), ordered = T)
island$MALE <- factor(island$MALE)
island$TREATMENT <- factor(island$TREATMENT,
                           levels = c("CONTROL", "CALM", "LOUD"), ordered = T)
island <- as.data.table(island)

# ------------ graphical_analysis ------------
# summary of response
response_summary <- with(island, summary(TIME))
# distribution of the response
response_hist <- ggplot(island, aes(x = TIME, y = ..density..)) +
  geom_bar(binwidth = 5, colour = "black", fill = NA) +
  labs(y = "DENSITY", title = "HISTOGRAM OF THE RESPONSE") +
  geom_density(size = 1, adjust = 0.7)
response_hist

# main effects
# over gender
gender_box <- ggplot(island, aes(x = MALE, y = TIME)) + geom_boxplot() +
  scale_x_discrete(limit = c("TRUE", "FALSE"), label = c("MALE", "FEMALE")) +
  labs(x = "GENDER", title = "RESPONSE OVER GENDER")
gender_box

# over age
age_box <- ggplot(island, aes(x = AGE_GROUP, y = TIME)) + geom_boxplot() +
  labs(x = "AGE GROUP", title = "RESPONSE OVER AGE GROUP")
age_box

# over treatment
treatment_box <- ggplot(island, aes(x = TREATMENT, y = TIME)) + geom_boxplot() +
  labs(x = "TREATMENT", title = "RESPONSE OVER TREATMENT")
treatment_box

# interaction with gender
int_male_data <- island[, .(GROUP_MEAN = mean(TIME)), by = .(TREATMENT, MALE)]
int_male_plot <- ggplot(int_male_data, aes(x = TREATMENT, y = GROUP_MEAN)) +
  geom_line(aes(colour = MALE, linetype = MALE, group = MALE)) +
  labs(y = "TIME", colour = "GENDER", linetype = "GENDER",
       title = "INTERACTIONS WITH GENDER") +
  scale_linetype_discrete(limits = c("TRUE", "FALSE"),
```

```
                          label = c("MALE", "FEMALE")) +
  scale_colour_discrete(limits = c("TRUE", "FALSE"),
                        label = c("MALE", "FEMALE"))
int_male_plot

# interaction with age
int_age_data <- island[, .(GROUP_MEAN = mean(TIME)), by = .(TREATMENT, AGE_GROUP)]
int_age_plot <- ggplot(int_age_data, aes(x = TREATMENT, y = GROUP_MEAN)) +
  geom_line(aes(colour = AGE_GROUP, linetype = AGE_GROUP, group = AGE_GROUP)) +
  labs(y = "TIME", colour = "AGE GROUP", linetype = "AGE GROUP",
       title = "INTERACTIONS WITH AGE GROUP")
int_age_plot

# ------------- anova -----------
# run anova
island_aov <- aov(TIME ~ MALE * TREATMENT + AGE_GROUP * TREATMENT +
                    MALE : AGE_GROUP + Error(ID / TREATMENT), island)
# take a look
summary(island_aov)
# see the effects
effects_table <- model.tables(island_aov)
effects_table

# sphericity test
# create a format it wide
island_wide <- dcast(island, ... ~ TREATMENT, value.var = "TIME")
dvm <- with(island_wide, cbind(CONTROL, CALM, LOUD))
idata <- with(island, data.frame(TREATMENT = factor(levels(TREATMENT))))
mlm <- lm(dvm ~ 1)
anova_car <- Anova(mlm, idata = idata, idesign = ~ TREATMENT, type = 3)
summary(anova_car)

# confidence intervals for main effects
# generate ci for treatment
setkey(island, TREATMENT)
con_cal <- t.test(island["CONTROL", TIME], island["CALM", TIME], paired = T,
                  var.equal = T, conf.level = 0.99)
con_loud <- t.test(island["CONTROL", TIME], island["LOUD", TIME], paired = T,
                   var.equal = T, conf.level = 0.99)
cal_loud <- t.test(island["CALM", TIME], island["LOUD", TIME], paired = T,
                   var.equal = T, conf.level = 0.99)

# create aggregate data for each subject
id_aggregate <- island[, .(AGE_GROUP = AGE_GROUP[1], MALE = MALE[1],
                           ID_MEAN = mean(TIME)), by = ID]
```

```
# ci for gender
setkey(id_aggregate, MALE)
male_female <- t.test(id_aggregate[MALE == T, ID_MEAN],
                      id_aggregate[MALE == F, ID_MEAN],
                      var.equal = F, paired = F, conf.level = 0.99)

# ci for age
setkey(id_aggregate, AGE_GROUP)
young_medium <- t.test(id_aggregate["YOUNG", ID_MEAN],
                       id_aggregate["MEDIUM", ID_MEAN],
                       paired = F, var.equal = F, conf.level = 0.99)
young_old <- t.test(id_aggregate["YOUNG", ID_MEAN],
                    id_aggregate["OLD", ID_MEAN],
                    paired = F, var.equal = F, conf.level = 0.99)
medium_old <- t.test(id_aggregate["MEDIUM", ID_MEAN],
                     id_aggregate["OLD", ID_MEAN],
                     paired = F, var.equal = F, conf.level = 0.99)
```