

Final Project

PSTAT122: Design and Analysis of Experiments

Fall 2025

STUDENT NAME

- Alex Dieter (A070X06)
- Ayden Kelleher (7874209)
- Darren Tran (A1H9J10)
- Peyton Lynch (A0V9Y96)

Due Date

Due Date: Monday, December 8, 2025, 11:59 PM

1 Introduction

Reaction times are important because they offer a clear window into how efficiently the brain and body process information and respond to changing conditions. In everyday life, quick reaction times contribute to safety, whether it's braking to avoid an accident, adjusting your balance to prevent a fall, or responding to unexpected changes in your environment. In sports and high pressure tasks, faster reaction times can be the difference between success and failure, making them central to training and evaluation. Ultimately, reaction times matter because they reflect how well we perceive, decide, and act in real time.

However, the factors that influence ones reaction times are not very well known. What we are looking to find are what influences someones reaction time. What we intend to discover are the things that influence someones reaction time with common factors, namely whether or not you are listening to music, whether or not the lights are on or off, and whether you are using your dominant or non dominant hand. We will figure out which of these factors truly influence reaction time, and if there is any interaction.

2 Experimental Design

Our experiment will be a 2^3 factorial experiment with blocks on what impacts reaction times, and how those factors interact. Our 3 factors will be the whether or not lights are on or off, whether or not you are listening to music, and whether or not you are using your dominant or non dominant hand. The way we will apply these treatments is by running our tests while the lights are on/off, whether or not the person being tested is listening to music, and have the person use their dominant or non dominant hand.

We are measuring reaction time, through the following website, <https://humanbenchmark.com/tests/reactiontime>. The way this website works is the screen changes color and you press the mouse as fast as you can after the screen color changes, and it records the time it takes to click. For each click you get one recorded time, so for each set of factors, we will do 5 tests and then record the mean. We will conduct the whole test once, recording each individual reaction test as an experimental unit, so we have a better idea of the variance of the data. After some quick preliminary testing while choosing which website to use, we found that it takes less than 30 seconds to run through 5 experimental units, which means our testing should be quick, in accordance to the project guidelines.

All of our factors will be fixed, since each factor contains each of the possible factors in that category. There is not a population of factors we are drawing from. Lights can be on or off, you can be using your dominant or non-dominant hand, and the tests can be conducted while the person is and isn't listening to music.

First, we will randomly order each of the treatments using code snippet 1 (7.1). The way this code works is it takes a random sample without replacement, which effectively gives us a random ordering of the tests. We will do this for each person, who will each be a block.

The way our experiment will use replication is very simple. The website we are using to track our reaction times administers 5 tests and then returns the mean. Our experimental unit will be the time it takes for one individual click. This means each test we get 5 replications per factor. We will do these replications back to back.

Our blocks will be different people. Our group consists of 4 people, so we will each be a block. Thus for each combination of treatments and blocks, we get 5 replicates.

- **Sample size:** Provide number of observations per condition. Guideline: 5–10 per treatment for CRD, 3–5 blocks for RCBD, total feasible within 1 hour.

3 Data Collection

We collected all of our data on the same day, with each of us doing the experiment in our own rooms. Everyone used their own laptop and mouse, but we made sure the room was quiet before starting. To get used to the Human Benchmark test, we each did five practice clicks in a normal setting with no factors applied.

After the warm-up, we followed the randomized order from Snippet 1. For every combination of music on/off, lights on/off, and dominant/weak hand, we ran Human Benchmark five separate times. Each run gave us one reaction time in milliseconds, so we ended up with five observations for each treatment combination.

We didn't run into any errors or bad trials, so we didn't have to repeat anything. Once we finished all of the runs, we entered the reaction times into R using the data-entry function in the Appendix. Each row of the dataset represents one run, along with the subject, the factor levels, and the replicate number. This gave us a full dataset with 160 observations that we used for the exploratory plots and ANOVA.

4 Analysis

- **Hypothesis Testing:** Test your hypothesis with an appropriate statistical test (e.g., ANOVA).
- **Tables, Figures, & Code:** Include key results (ANOVA table, post-test) and relevant R code excerpts where needed.
- Use R to analyze the data.

5 Conclusions

- Summarize key findings.
- Comment on limitations and possible improvements.

6 References

<https://humanbenchmark.com/tests/reactiontime>

Kandethody M. Ramachandran, Chris P. Tsokos, Chapter 9 - Analysis of variance, Editor(s): Kandethody M. Ramachandran, Chris P. Tsokos, Mathematical Statistics with Applications in R (Third Edition), Academic Press, 2021, Pages 369-414, ISBN 9780128178157, <https://doi.org/10.1016/B978-0-12-817815-7.00009-9>. (<https://www.sciencedirect.com/science/article/pii/B9780128178157000099>)

7 Appendices

- R code.

7.1 Snippet 1

```
1 set.seed(12032025)
2 comboOfFactors <- c('Quiet&Dark&DomHand', 'Music&Dark&DomHand',
3                     'Quiet&Dark&WeakHand', 'Music&Dark&WeakHand',
4                     'Quiet&Light&DomHand', 'Music&Light&DomHand',
5                     'Quiet&Light&WeakHand', 'Music&Light&WeakHand')
6 Person1 <- sample(comboOfFactors, size = 8)
7 Person2 <- sample(comboOfFactors, size = 8)
8 Person3 <- sample(comboOfFactors, size = 8)
9 Person4 <- sample(comboOfFactors, size = 8)
10
11 cat('Alex will test in the following order\n', Person1[c(1,2,3,4)], '\n', Person1[c(5,6,7,8)], '\n')
```

Alex will test in the following order

Music&Light&WeakHand Quiet&Dark&WeakHand Quiet&Light&DomHand Music&Dark&WeakHand
Music&Light&DomHand Music&Dark&DomHand Quiet&Light&WeakHand Quiet&Dark&DomHand

```
1 cat('\nAyden will test in the following order\n', Person2[c(1,2,3,4)], '\n', Person2[c(5,6,7,8)], '\n')
```

Ayden will test in the following order

Quiet&Dark&WeakHand Music&Dark&WeakHand Music&Light&DomHand Quiet&Light&WeakHand
Music&Light&WeakHand Quiet&Light&DomHand Quiet&Dark&DomHand Music&Dark&DomHand

```
1 cat('\nDarren will test in the following order\n', Person3[c(1,2,3,4)], '\n', Person3[c(5,6,7,8)], '\n')
```

Darren will test in the following order

Quiet&Dark&WeakHand Music&Dark&WeakHand Music&Light&DomHand Music&Light&WeakHand
Quiet&Light&WeakHand Quiet&Dark&DomHand Music&Dark&DomHand Quiet&Light&DomHand

```
1 cat('\nPeyton will test in the following order\n', Person4[c(1,2,3,4)], '\n', Person4[c(5,6,7,8)], '\n')
```

Peyton will test in the following order

Music&Light&WeakHand Music&Light&DomHand Quiet&Dark&WeakHand Quiet&Light&DomHand
Music&Dark&DomHand Quiet&Dark&DomHand Music&Dark&WeakHand Quiet&Light&WeakHand

```
1 library(dplyr)
2
3 subjects      <- c("Alex", "Ayden", "Darren", "Peyton")
4 music_levels  <- c("quiet", "music")
5 lights_lvls   <- c("dark", "light")
6 hand_levels   <- c("dom", "weak")
7
8 # 3-factor 2x2x2 × 4 subjects × 5 reps
9 df <- expand.grid(
10   subject = subjects,
11   music   = music_levels,
12   lights  = lights_lvls,
13   hand    = hand_levels,
14   rep     = 1:5          # 5 runs per combo
15 ) %>%
```

```

16   arrange(subject)
17
18   # column where you'll put the reaction times
19   df$time_ms <- NA_real_

```

```

1   #This is the function so we can enter in all of our data easily
2   enter_times <- function(name, music, lights, hand, times) {
3     df$time_ms[df$subject==name &
4               df$music==music &
5               df$lights==lights &
6               df$hand==hand] <- times
7   }
8
9   # Alex's Data
10  enter_times("Alex", "music", "light", "weak", c(265, 217, 219, 261, 260))
11  enter_times("Alex", "quiet", "dark", "weak", c(250, 271, 259, 202, 216))
12  enter_times("Alex", "quiet", "light", "dom", c(207, 206, 203, 126, 189))
13  enter_times("Alex", "music", "dark", "weak", c(181, 236, 194, 215, 214))
14  enter_times("Alex", "music", "light", "dom", c(197, 189, 97, 200, 181))
15  enter_times("Alex", "music", "dark", "dom", c(197, 147, 96, 194, 201))
16  enter_times("Alex", "quiet", "light", "weak", c(214, 183, 241, 87, 221))
17  enter_times("Alex", "quiet", "dark", "dom", c(212, 202, 203, 187, 159))
18
19  # Ayden's Data
20  enter_times("Ayden", "quiet", "dark", "weak", c(223, 245, 235, 226, 236))
21  enter_times("Ayden", "music", "dark", "weak", c(226, 233, 229, 235, 233))
22  enter_times("Ayden", "music", "light", "dom", c(220, 231, 275, 226, 238))
23  enter_times("Ayden", "quiet", "light", "weak", c(252, 219, 218, 243, 235))
24  enter_times("Ayden", "music", "light", "weak", c(226, 255, 247, 228, 242))
25  enter_times("Ayden", "quiet", "light", "dom", c(226, 338, 227, 196, 245))
26  enter_times("Ayden", "quiet", "dark", "dom", c(224, 232, 223, 230, 231))
27  enter_times("Ayden", "music", "dark", "dom", c(220, 241, 238, 257, 234))
28
29  # Darren's Data
30  enter_times("Darren", "quiet", "dark", "weak", c(269, 264, 272, 274, 253))
31  enter_times("Darren", "music", "dark", "weak", c(130, 290, 273, 289, 291))
32  enter_times("Darren", "music", "light", "dom", c(247, 230, 118, 260, 232))
33  enter_times("Darren", "music", "light", "weak", c(228, 251, 270, 261, 275))
34  enter_times("Darren", "quiet", "light", "weak", c(290, 268, 260, 245, 271))
35  enter_times("Darren", "quiet", "dark", "dom", c(267, 265, 248, 231, 259))
36  enter_times("Darren", "music", "dark", "dom", c(271, 293, 307, 260, 275))
37  enter_times("Darren", "quiet", "light", "dom", c(223, 215, 221, 229, 253))
38
39  # Peyton's Data
40  enter_times("Peyton", "music", "light", "weak", c(272, 265, 279, 236, 312))
41  enter_times("Peyton", "music", "light", "dom", c(287, 243, 265, 229, 259))
42  enter_times("Peyton", "quiet", "dark", "weak", c(222, 329, 233, 289, 278))
43  enter_times("Peyton", "quiet", "light", "dom", c(228, 249, 295, 264, 232))
44  enter_times("Peyton", "music", "dark", "dom", c(252, 229, 243, 250, 239))
45  enter_times("Peyton", "quiet", "dark", "dom", c(223, 226, 238, 218, 217))
46  enter_times("Peyton", "music", "dark", "weak", c(228, 243, 291, 220, 312))
47  enter_times("Peyton", "quiet", "light", "weak", c(260, 261, 275, 249, 258))

```

After collecting all of the reaction time measurements and entering them into R, we summarized the data by computing the mean and standard deviation for each of the eight treatment combinations (Music \times Lights \times Hand). This gives a quick overview of how reaction times behaved across the different conditions. The table below shows the average reaction time (in milliseconds) and the variability across the five replicates for each treatment:

```

1 library(dplyr)
2 library(knitr)
3
4 summary_table <- df %>%
5   dplyr::group_by(music, lights, hand) %>%
6   dplyr::summarise(
7     mean_time = mean(time_ms),
8     sd_time   = sd(time_ms)
9   )
10
11 kable(summary_table,
12       col.names = c("Music", "Lights", "Hand", "Mean Time (ms)", "Std Dev (ms)"),
13       digits = 4,
14       caption = "Summary of Reaction Times by Treatment Group",
15       align = 'c')

```

Table 1: Summary of Reaction Times by Treatment Group

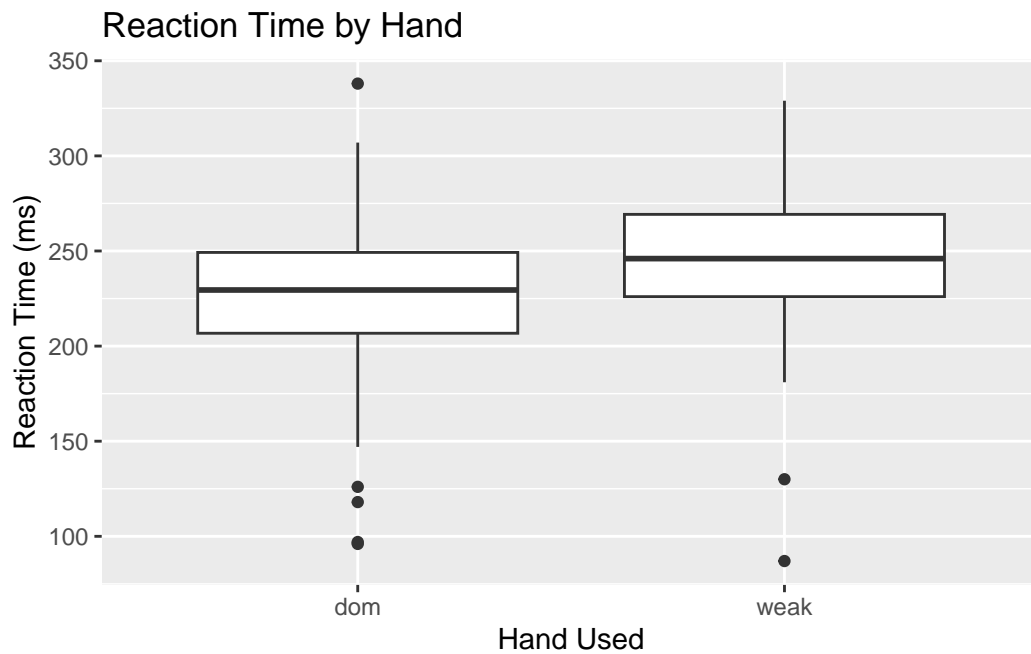
Music	Lights	Hand	Mean Time (ms)	Std Dev (ms)
quiet	dark	dom	224.75	25.6163
quiet	dark	weak	252.30	29.7995
quiet	light	dom	228.60	42.3002
quiet	light	weak	237.50	43.4166
music	dark	dom	232.20	48.4731
music	dark	weak	238.15	43.7183
music	light	dom	221.20	47.9249
music	light	weak	253.45	23.6609

Overall, reaction times varied quite a bit across conditions. In general, the weak hand tended to be slower than the dominant hand in most settings, which matches what we expected. Music and lighting also seemed to influence the results, but the direction of the effect wasn't immediately clear from the raw means alone. We use these summaries mainly to get a sense of the patterns in the data before moving on to the formal ANOVA in the analysis section.

```

1 library(ggplot2)
2
3 # Boxplot by Hand
4 ggplot(df, aes(x = hand, y = time_ms)) +
5   geom_boxplot() +
6   labs(title = "Reaction Time by Hand",
7        x = "Hand Used",
8        y = "Reaction Time (ms)")

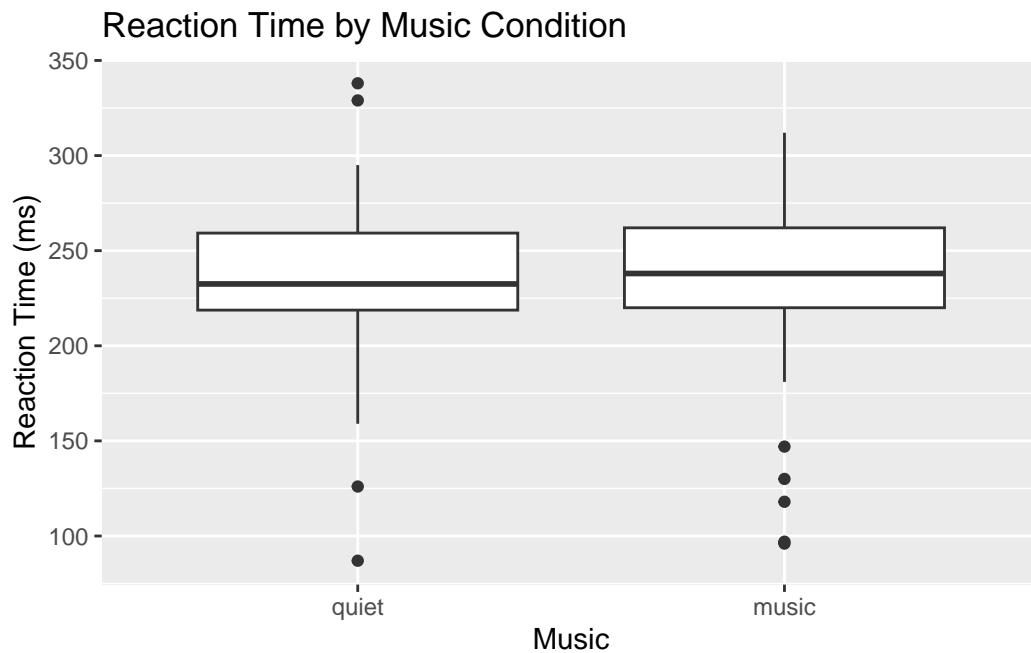
```



```

1 # Boxplot by Music
2 ggplot(df, aes(x = music, y = time_ms)) +
3   geom_boxplot() +
4   labs(title = "Reaction Time by Music Condition",
5         x = "Music",
6         y = "Reaction Time (ms)")

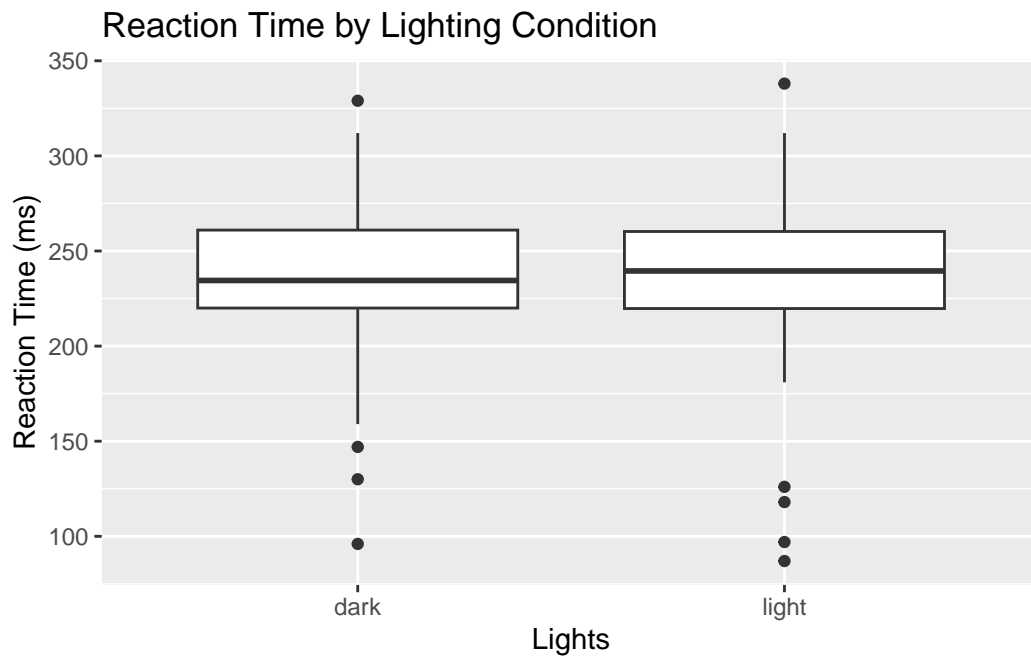
```



```

1 # Boxplot by Lights
2 ggplot(df, aes(x = lights, y = time_ms)) +
3   geom_boxplot() +
4   labs(title = "Reaction Time by Lighting Condition",
5         x = "Lights",
6         y = "Reaction Time (ms)")

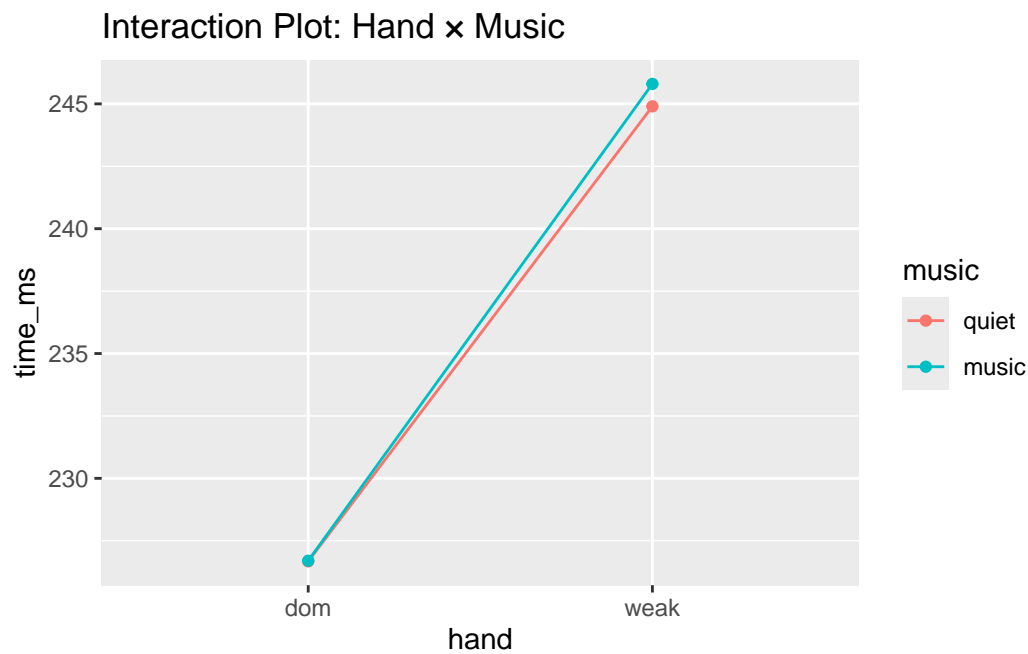
```



```

1 # Interaction Plot: Hand × Music
2 ggplot(df, aes(x = hand, y = time_ms, color = music, group = music)) +
3   stat_summary(fun = mean, geom = "line") +
4   stat_summary(fun = mean, geom = "point") +
5   labs(title = "Interaction Plot: Hand × Music")

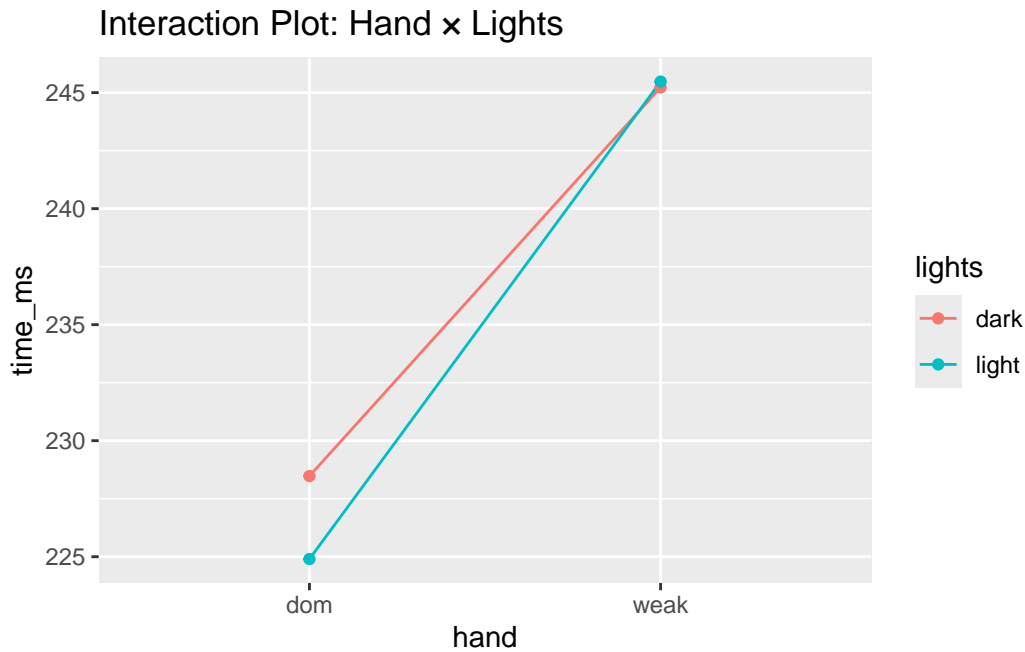
```



```

1 # Interaction Plot: Hand × Lights
2 ggplot(df, aes(x = hand, y = time_ms, color = lights, group = lights)) +
3   stat_summary(fun = mean, geom = "line") +
4   stat_summary(fun = mean, geom = "point") +
5   labs(title = "Interaction Plot: Hand × Lights")

```



7.2 Exploratory Data

To get a first look at our data, we made several boxplots for each factor and a set of interaction plots. The boxplot for hand (dominant vs. weak) showed the clearest difference. The weak hand had noticeably slower reaction times on average, and the spread was larger. This matched what we expected, since most people are less coordinated with their non-dominant hand.

For the music factor, the boxplots for “quiet” and “music” looked fairly similar overall. The medians were close, and both conditions had a few outliers. There might still be a small effect, but it wasn’t obvious just from the boxplots. The lighting factor (dark vs. light) showed a similar pattern. There were slight differences between the medians, but nothing significant.

We also made interaction plots to check whether the factors influenced each other. The hand × music plot showed two lines that were almost parallel, meaning music probably doesn’t change the effect of hand very much. The hand × lights plot looked similar, with only small differences between the light and dark conditions. All of the lines went up from dominant to weak hand, which lines up with the boxplots.

Overall, the exploratory graphs suggested that the hand factor has the strongest effect, while music and lighting might have smaller or weaker effects. The interaction plots didn’t show any major crossings, so any interactions are likely to be small. These patterns helped us get a feel for the data before running the formal ANOVA.

7.3 Analysis

Now that we have a basic idea of what our data looks like, we can run an ANOVA test to see which factors have a significant effect on the response variable.

```
1 anova_model <- aov(time_ms ~ hand*lights*music+subject, data=df)
2 summary(anova_model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
hand	1	13932	13932	13.283	0.000369	***
lights	1	111	111	0.105	0.745888	
music	1	9	9	0.008	0.928153	
subject	3	78110	26037	24.824	4.39e-13	***
hand:lights	1	146	146	0.139	0.709315	
hand:music	1	8	8	0.007	0.932027	
lights:music	1	581	581	0.554	0.457724	


```

hand:lights:music    1    5051    5051    4.816 0.029744 *
Residuals           149 156276    1049
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

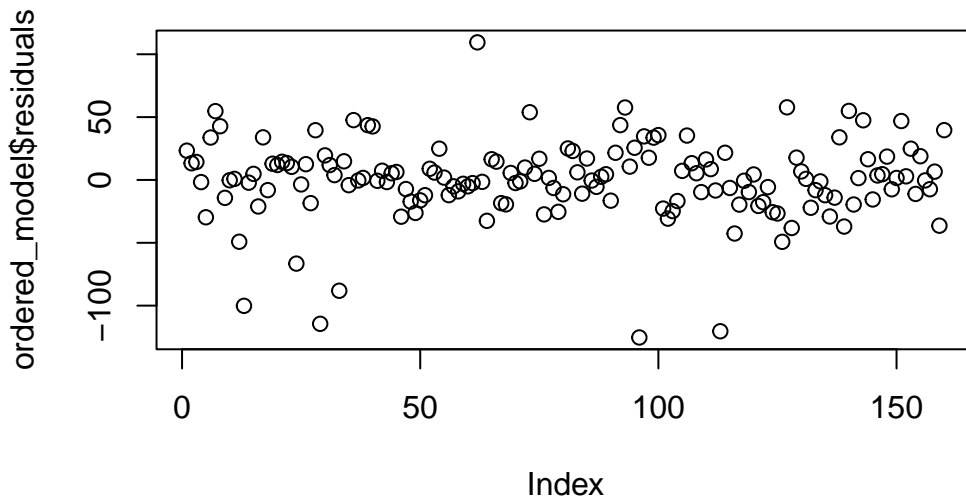
Recall the assumptions of the ANOVA test, that the trials are independent, that the residuals are normally distributed, and that the residuals have equal variance.

Our experimental design should ensure that the trials were independent. To double check, we will plot the residuals in time order, to see if there is a strong correlation between residuals. A tendency to have runs of positive and negative residuals indicates positive correlation.

```

1 df_ordered <- df %>%
2   mutate(hand = factor(hand, levels = c("dom", "weak"))) %>% # Specify desired order
3   arrange(hand) %>%
4   mutate(music = factor(music, levels = c('quiet', 'music'))) %>%
5   arrange(music) %>%
6   mutate(lights = factor(lights, levels = c('dark', 'light'))) %>%
7   arrange(lights) %>%
8   mutate(subject = factor(subject, levels = c('Alex', 'Ayden', 'Darren', 'Peyton'))) %>%
9   arrange(subject)
10
11 #df_ordered
12 ordered_model <- aov(time_ms ~ hand*lights*music+subject, data=df_ordered)
13
14 plot(ordered_model$residuals)

```

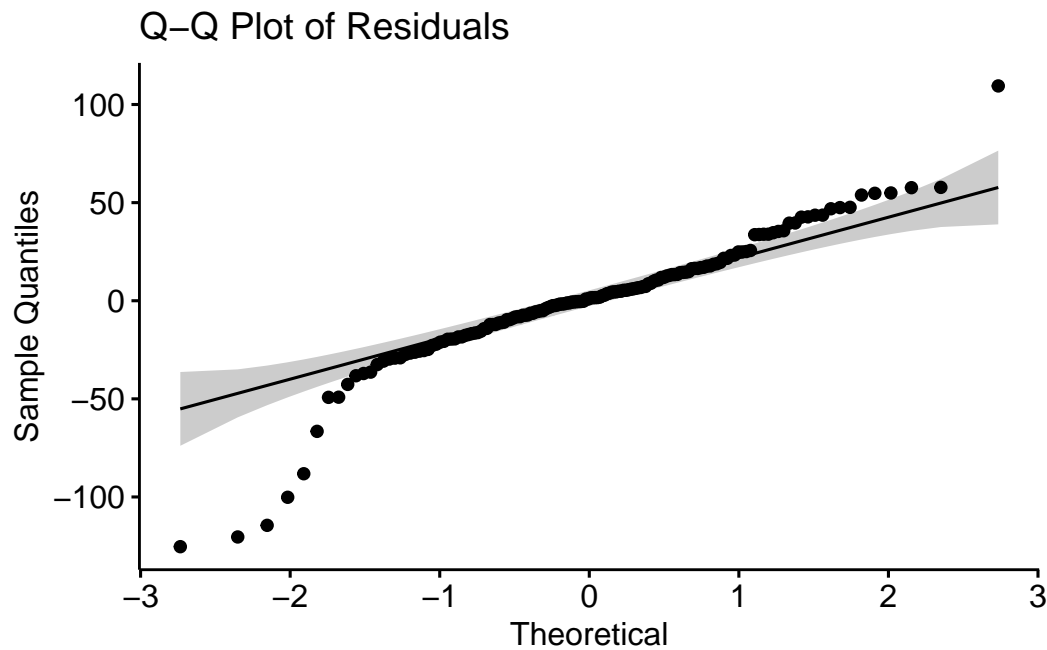


After plotting the ordered residuals, we can see that there does not appear to be strong correlation between subsequent residuals, thus we can confirm our trials are independent.

```

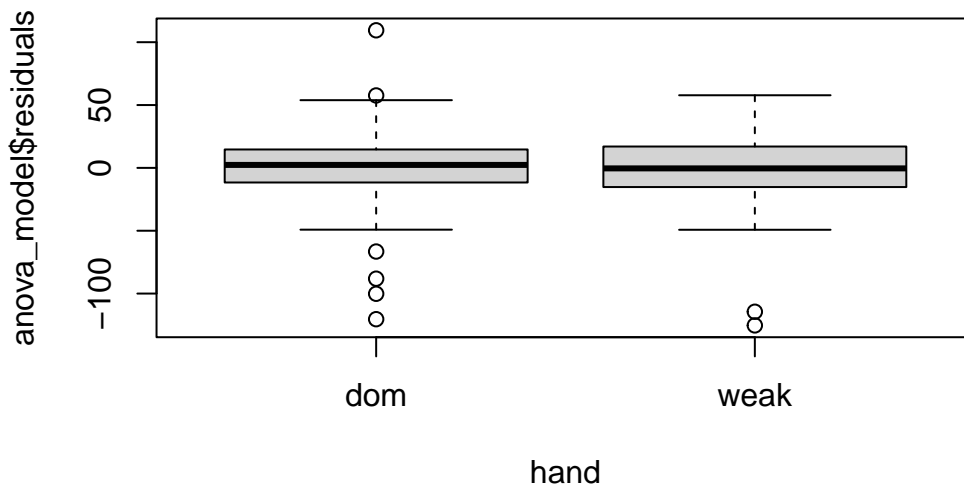
1 library(ggpubr)
2 library(knitr)
3
4 anova_resid <- residuals(anova_model)
5 ggqqplot(anova_resid, title = "Q-Q Plot of Residuals", ylab = "Sample Quantiles")

```

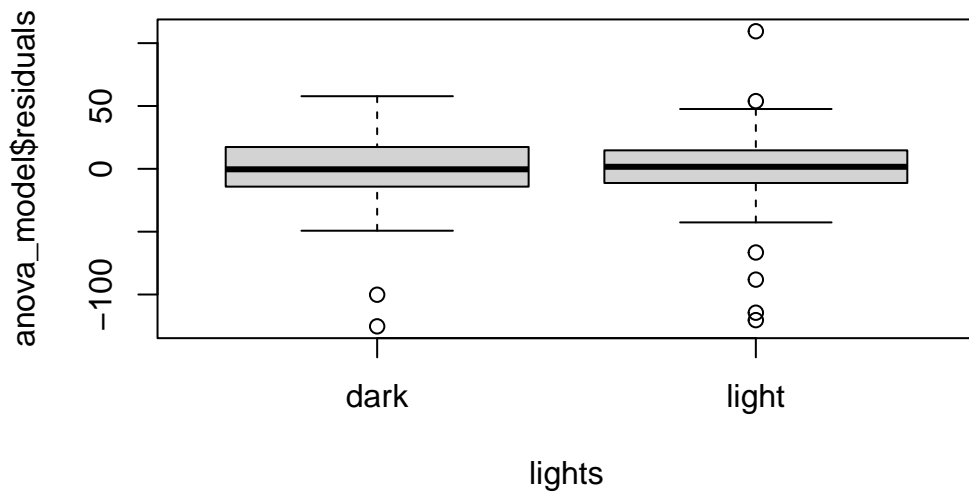


To check the assumption of normality for the ANOVA model we have, we will use the Q-Q plot of the residuals, which allows us to see if the residuals are distributed normally visually. From the Q-Q plot it is seen that the points follow the line closely and remain the 95% confidence band. We have some outliers in our tails, with larger than expected residuals. However, the normality assumption is robust for the fixed effect anova test, which means moderate departures from normality are of little concern. Thus we can say our residuals are normally distributed or close enough to be of little concern.

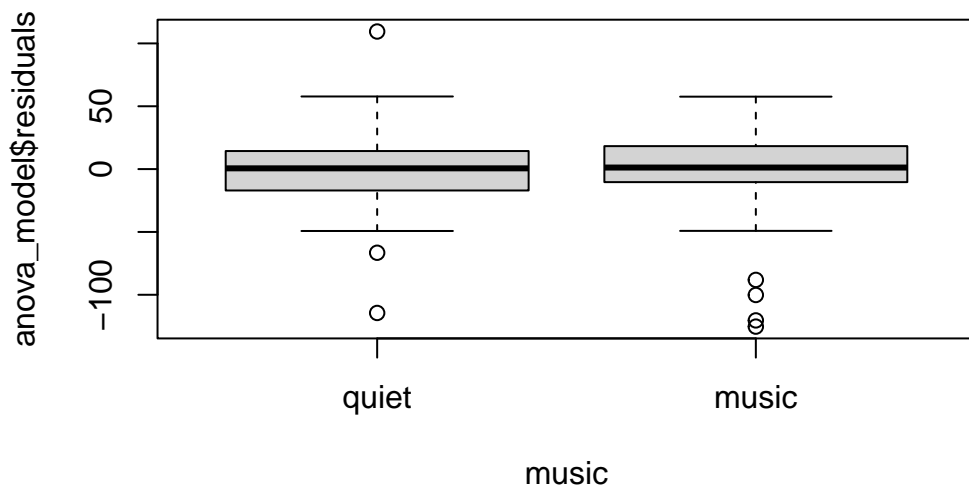
```
1 plot(anova_model$residuals ~ hand, data=df)
```



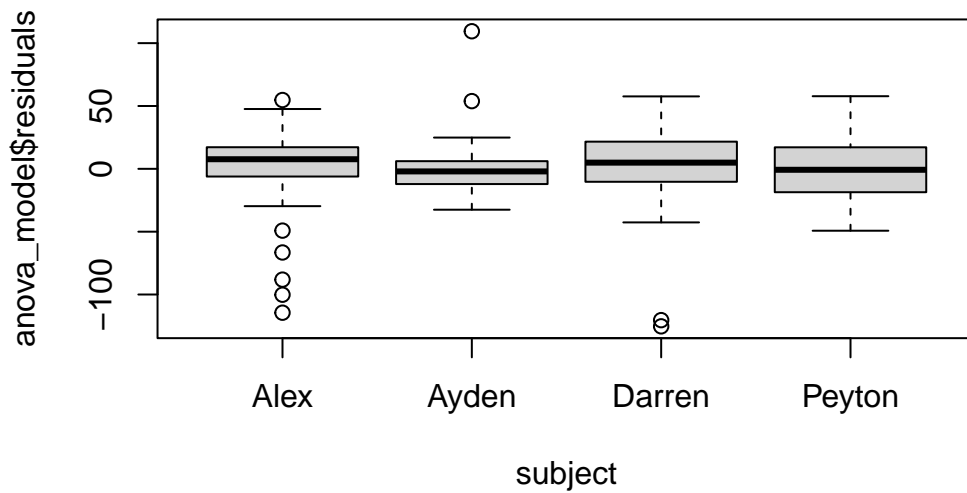
```
1 plot(anova_model$residuals ~ lights, data=df)
```



```
1 plot(anova_model$residuals ~ music, data=df)
```



```
1 plot(anova_model$residuals ~ subject, data=df)
```



Our last assumption to check is equal variance, or homoscedasticity. To test it, we create a box plot of residuals for each factor. Looking at the boxplot, we can see that the variances for each main effect are about equal. However, the variance does seem to depend on the blocks. To verify this, we will use the bartlett test

```
1 subject_variance <- bartlett.test(anova_model$residuals ~ subject, data=df)
2 cat( 'The p-value that all subjects have the same variance is',
3     subject_variance$p.value
4 )
```

The p-value that all subjects have the same variance is 0.008917946

```
1 sd_df <- data.frame(
2   subject = c('Alex','Ayden','Darren','Peyton'),
3   sample_variance = c(
4     var(anova_model$residuals[df$subject=='Alex']),
5     var(anova_model$residuals[df$subject=='Ayden']),
6     var(anova_model$residuals[df$subject=='Darren']),
7     var(anova_model$residuals[df$subject=='Peyton'])
8   )
9 )
10
11
12 kable(
13   sd_df,
14   digits=4,
15   row.names=F
16 )
```

subject	sample_variance
Alex	1462.8298
Ayden	579.8910
Darren	1274.6288

subject	sample_variance
Peyton	689.7243

We can see that the p-value for the test between different subjects is 0.008, which is significant for $\alpha = .05$. Thus we can conclude that the variance is different for different subjects. However, we have equal sample sizes for each block and we have similar sized variances, where the largest is less than double than the smallest. According to chapter 9.3.2 of Ramachandran and Tsokos, “If the sizes of all the samples are equal, ANOVA is mostly robust for violation of homogeneity of the variances. A rule of thumb used for robustness for this condition is that the ratio of sample variance of the largest sample variance s^2 to the smallest sample variance s^2 should be no more than 3:1”. We can see that our sample variances are under this limit, so we can say our anova test will be robust against this unequal variance.

Therefore, we can say that fulfills all the assumptions.

Recall that the null hypothesis test for an ANOVA test is all the means equal each other, and the alternate hypothesis is that at least one mean is different. After running our ANOVA test, we can see that with significance level $\alpha = .05$, we have which hand you use is significant with a p-value of 0.000369, the interaction between all three is significant with p-value of 0.029744, and the subjects have a significant effect with a p-value of 4.39e-13. However, if we use a Bonferroni adjusted α value, we get $\alpha_{Bonferroni} = \alpha/7 = .05/7 = .0071$, which means that only the hand used and subject (our blocks) are significant.

```

1 library(knitr)
2 library(DescTools)
3
4 bonferroni_test <- PostHocTest(anova_model, method = c("bonferroni"), conf.level = .05)
5
6 kable(
7   rbind(bonferroni_test$hand,
8         bonferroni_test$lights,
9         bonferroni_test$music,
10        bonferroni_test$subject),
11   digits = 4,
12   caption = "Bonferroni Post-Hoc Test Results",
13   align = 'c'
14 )

```

Table 3: Bonferroni Post-Hoc Test Results

	diff	lwr.ci	upr.ci	pval
weak-dom	18.6625	18.3409	18.9841	0.0004
light-dark	-1.6625	-1.9841	-1.3409	0.7459
music-quiet	0.4625	0.1409	0.7841	0.9282
Ayden-Alex	35.9750	25.7077	46.2423	0.0000
Darren-Alex	53.2250	42.9577	63.4923	0.0000
Peyton-Alex	54.9750	44.7077	65.2423	0.0000
Darren-Ayden	17.2500	6.9827	27.5173	0.1109
Peyton-Ayden	19.0000	8.7327	29.2673	0.0576
Peyton-Darren	1.7500	-8.5173	12.0173	1.0000

What this means is that we have evidence that whether or not you use your dominant hand is significant, different people have significant differences, and that we have poor evidence that the 3 way interaction term is significant.

7.4 Technical Limitations and Improvements

The nature of our experiment presents a couple of limitations and technical challenges. Within our blocking factor, everyone possessed different levels of hardware, specifically device differences. One person's laptop or mouse may have more or less input lag, which in turn influences reaction times. The use of a browser based test plays a role as well, as the technology introduces a level of delay that is not reflective of the subject's actual reaction time. Aside from instrumentation influences, the block effect also captures the effect of physiological differences. Not every human is built with the same reaction speed; this can be due to factors like age and overall health. Furthermore, outside factors like caffeine usage, stress, and level of rest can impact times significantly. Lastly, we utilized a straightforward approach to the music factor: quiet or with music. This binary "yes or no" approach does not account for the genre of music the subject may be listening to. The genre may impact times, as a subject could be listening to music with a significantly higher or lower tempo.

For future iterations of this experiment, it would be helpful to standardize the equipment, third-party influences, and the environment in which we conduct the test. Having everyone on the same setup, using the same laptop and mouse, would control the instrumentation noise that can impact the data. Honing in on the third-party influences that impact physiological performance, such as standardizing the use of stimulants like caffeine and the level of sleep, would help control the variation these factors introduce. Aside from standardizing, introducing a specific factor for stimulants would be another path to take. While the Human Benchmark website brings some documentation issues, the alternative plan having someone else manually time the person engaged in the experiment would introduce the issue of additional human error, as a timekeeper could easily stop the timer too early or too late. Ultimately the experiment of human reaction time, will always come with some form of uncontrollable human error.