

# Stats101B Project R Codes

Maya Gheewala mgheewala@ucla.edu      Jolina Hor jolinahor@g.ucla.edu  
Aashima Khanna ashkhanna06@g.ucla.edu      Ayushi Kadakia ayukadakia@ucla.edu  
Erica Song ericasong@ucla.edu

2025-05-29

## Read in Data

```
data <- read.csv("/Users/ayushikadakia/Downloads/STATS 101B Group Project Data - Ayushi.csv")
```

## Comparison of Chocolate Types by Time Level

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
data$time_level <- dplyr::recode(data$time_level,
                                "right_after" = "1_rightafter",
                                "one_hour" = "2_onehour",
                                "two_hours" = "3_twohour")

data$time_level <- factor(data$time_level, levels = c("1_rightafter", "2_onehour", "3_twohour"))

data$chocolate_level <- as.character(data$chocolate_level)

data$chocolate_level <- dplyr::recode(data$chocolate_level,
                                      "White" = "1_white",
                                      "40" = "2_40",
                                      "70" = "3_70",
                                      "85" = "4_85",
```

```

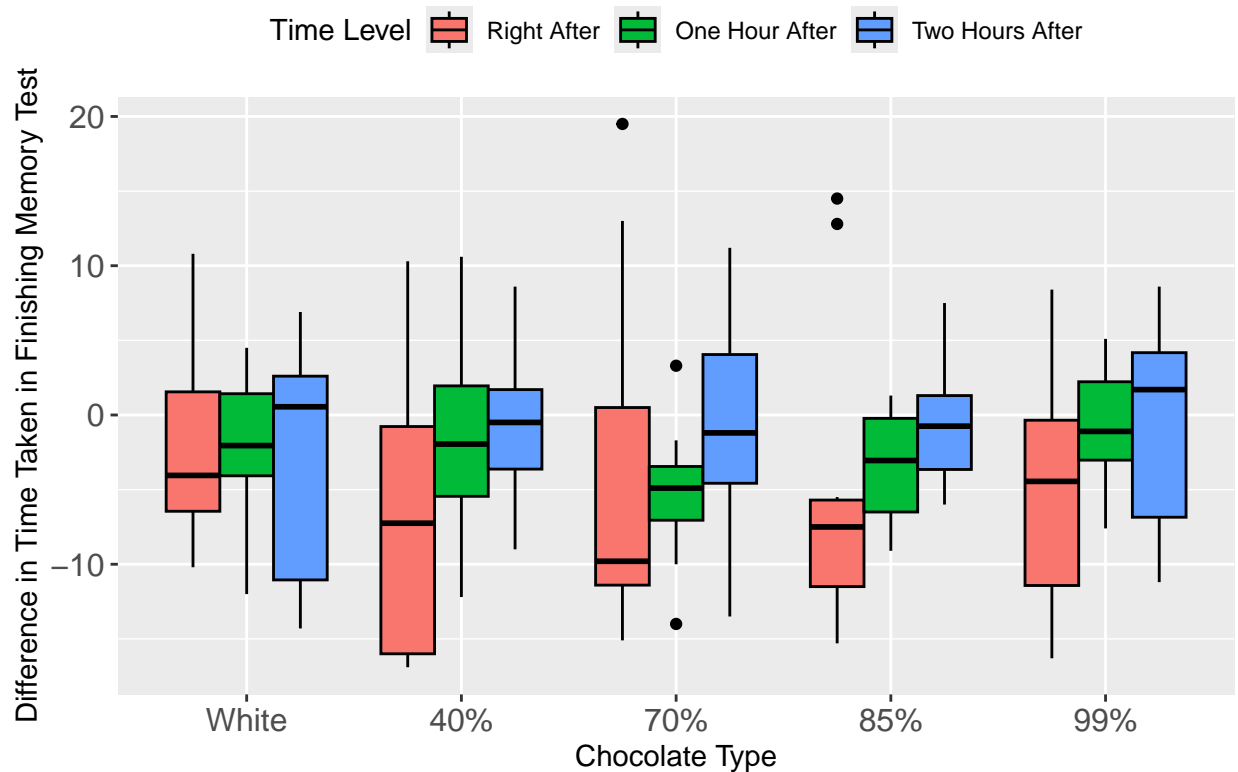
"99" = "5_99")

data$chocolate_level <- factor(data$chocolate_level,
                                levels = c("1_white", "2_40", "3_70", "4_85", "5_99"))

ggplot(data, aes(x = chocolate_level, y = difference, fill = time_level)) +
  geom_boxplot(position = position_dodge(width = 0.75), color = "black") +
  labs(
    title = "Comparison of Chocolate Types by Time Level",
    x = "Chocolate Type",
    y = "Difference in Time Taken in Finishing Memory Test",
    fill = "Time Level"
  ) +
  scale_fill_discrete(labels = c(
    "1_rightafter" = "Right After",
    "2_onehour" = "One Hour After",
    "3_twohour" = "Two Hours After"
  )) +
  scale_x_discrete(labels = c(
    "1_white" = "White",
    "2_40" = "40%",
    "3_70" = "70%",
    "4_85" = "85%",
    "5_99" = "99%"
  )) +
  theme(
    legend.position = "top",
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 16)
  )

```

## Comparison of Chocolate Types by Time Level



### Boxplot for Mean Difference based on Chocolate Level

```
library(dplyr)
library(ggplot2)

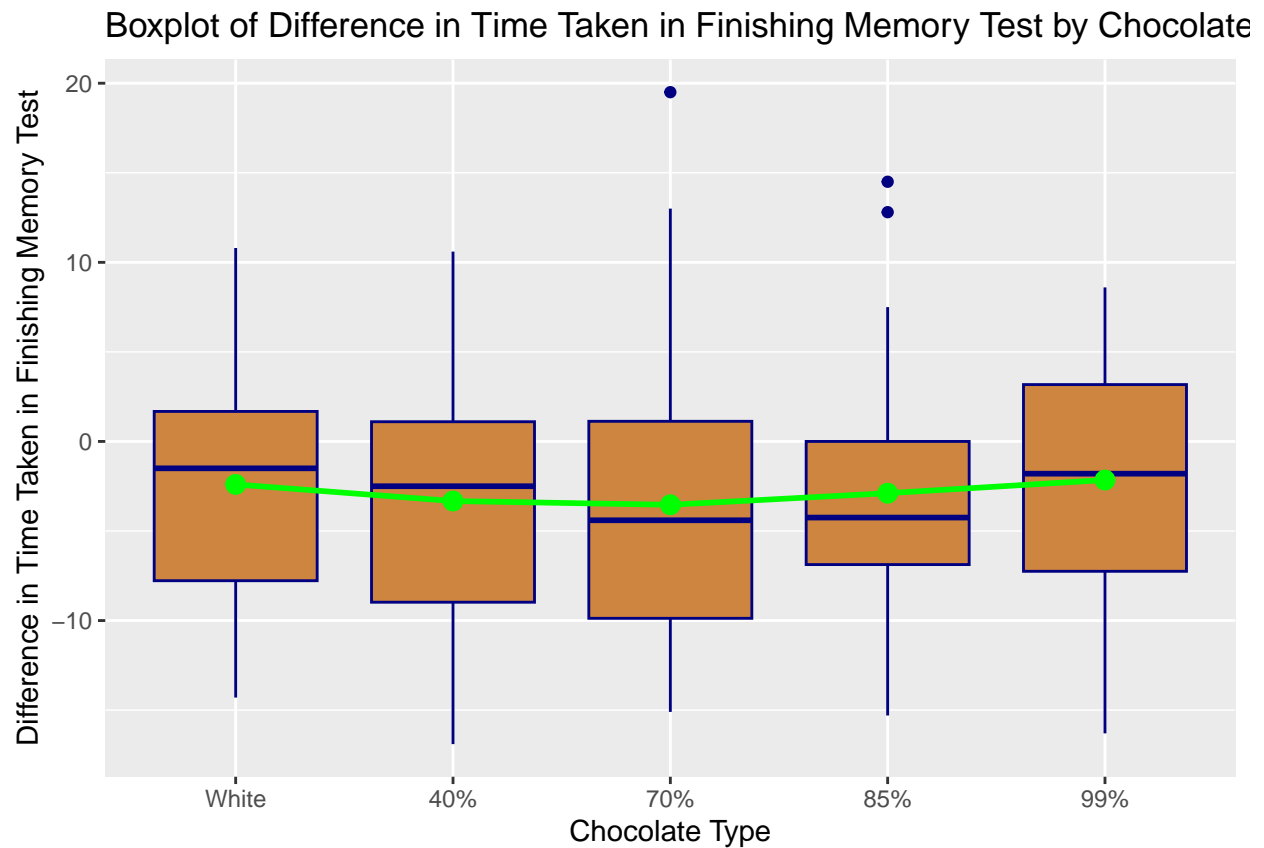
mean_data <- data %>%
  group_by(chocolate_level) %>%
  summarise(mean_diff = mean(difference, na.rm = TRUE))

ggplot(data = data, aes(x = factor(chocolate_level), y = difference)) +
  geom_boxplot(fill = "peru", color = "navy") +
  geom_point(data = mean_data, aes(x = factor(chocolate_level), y = mean_diff),
            color = "green", size = 3) +
  geom_line(data = mean_data, aes(x = factor(chocolate_level), y = mean_diff, group = 1),
            color = "green", linewidth = 1) +
  labs(
    title = "Boxplot of Difference in Time Taken in Finishing Memory Test by Chocolate Type",
    x = "Chocolate Type",
    y = "Difference in Time Taken in Finishing Memory Test"
  ) +
  scale_x_discrete(labels = c(
    "1_white" = "White",
    "2_40" = "40%",
    "3_70" = "70%",
```

```

"4_85" = "85%",
"5_99" = "99%"
))

```



## Comparison of Time Levels by Chocolate Type

```

library(ggplot2)
library(dplyr)

ggplot(data, aes(x = time_level, y = difference, fill = chocolate_level)) +
  geom_boxplot(position = position_dodge(width = 0.75), color = "black") +
  labs(
    title = "Comparison of Time Levels by Chocolate Type",
    x = "Time Level",
    y = "Difference in Time Taken in Finishing Memory Test",
    fill = "Chocolate Type"
  ) +
  scale_x_discrete(labels = c(
    "1_rightafter" = "Right After",
    "2_onehour" = "One Hour After",
    "3_twohour" = "Two Hours After"
  )) +
  scale_fill_discrete(labels = c(

```

```

"1_white" = "White",
"2_40" = "40%",
"3_70" = "70%",
"4_85" = "85%",
"5_99" = "99%"
)) +
theme(
  legend.position = "top",
  axis.text = element_text(size = 12),
  plot.title = element_text(hjust = 0.5, size = 16)
)

```



### Boxplot for Mean Difference based on Time

```

library(dplyr)

mean_data <- data %>%
  group_by(time_level) %>%
  summarise(mean_diff = mean(difference, na.rm = TRUE))

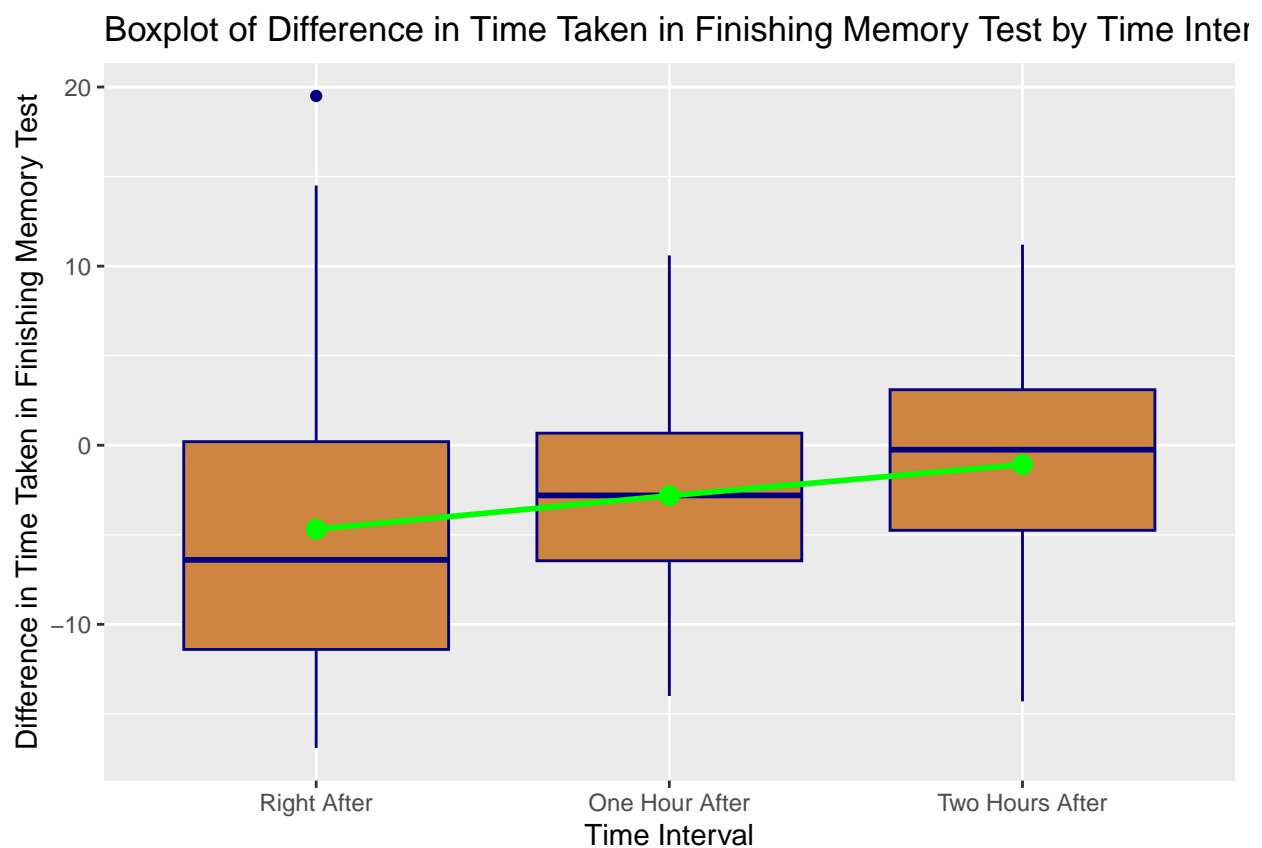
ggplot(data = data, aes(x = factor(time_level), y = difference)) +
  geom_boxplot(fill = "peru", color = "navy") +
  geom_point(data = mean_data, aes(x = factor(time_level), y = mean_diff),

```

```

    color = "green", size = 3) +
  geom_line(data = mean_data, aes(x = factor(time_level), y = mean_diff, group = 1),
    color = "green", linewidth = 1) +
  labs(
    title = "Boxplot of Difference in Time Taken in Finishing Memory Test by Time Interval",
    x = "Time Interval",
    y = "Difference in Time Taken in Finishing Memory Test"
  ) +
  scale_x_discrete(labels = c(
    "1_rightafter" = "Right After",
    "2_onehour" = "One Hour After",
    "3_twohour" = "Two Hours After"
  ))

```



## Interaction Plot

```

library(ggplot2)

data$chocolate_level <- as.factor(data$chocolate_level)
data$time_level <- as.factor(data$time_level)

interaction_means <- aggregate(difference ~ time_level + chocolate_level, data = data, FUN = mean)

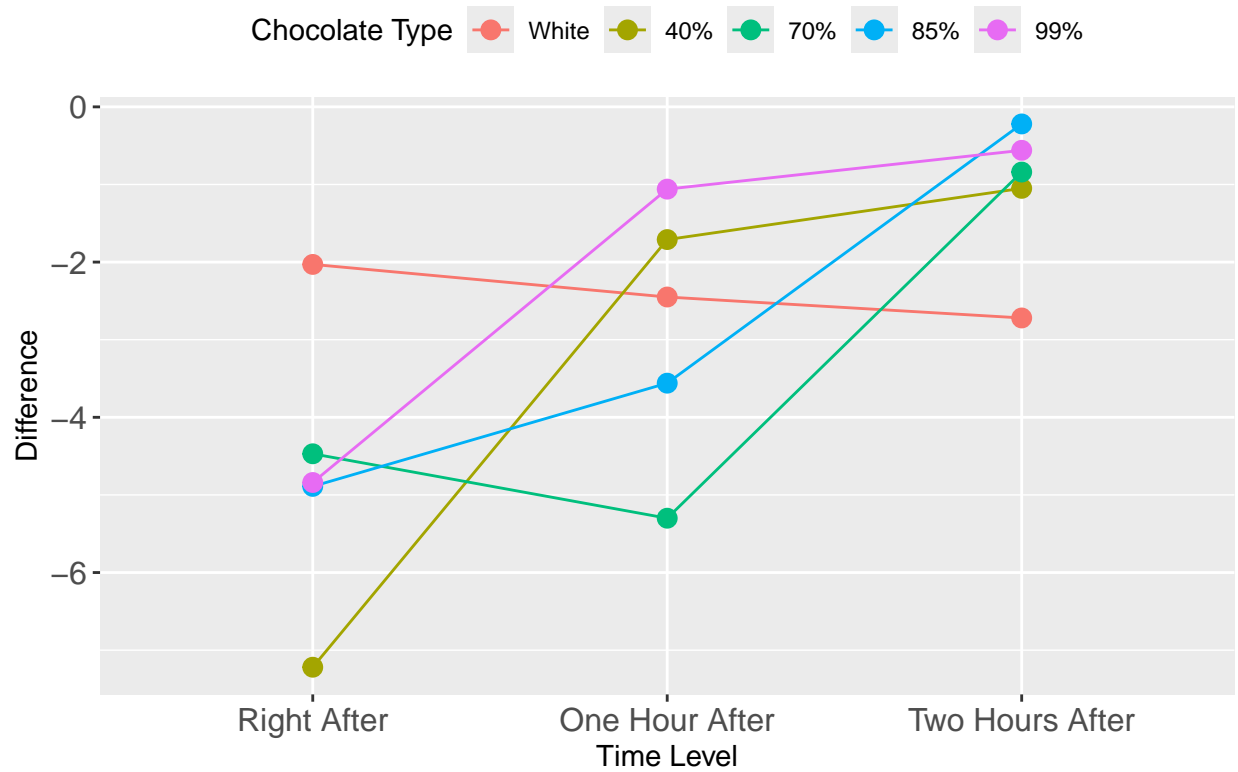
```

```

ggplot(interaction_means, aes(x = time_level, y = difference, group = chocolate_level, color = chocolate_level)) +
  geom_line() +
  geom_point(size = 3) +
  labs(
    title = "Interaction Plot",
    x = "Time Level",
    y = "Difference",
    color = "Chocolate Type"
  ) +
  scale_x_discrete(labels = c(
    "1_rightafter" = "Right After",
    "2_onehour" = "One Hour After",
    "3_twohour" = "Two Hours After"
  )) +
  scale_color_discrete(labels = c(
    "1_white" = "White",
    "2_40" = "40%",
    "3_70" = "70%",
    "4_85" = "85%",
    "5_99" = "99%"
  )) +
  theme(
    legend.position = "top",
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 16)
  )

```

## Interaction Plot



## ANOVA Summary

```
modell1 <- aov(difference ~ as.factor(chocolate_level) * as.factor(time_level) + as.factor(age_range),
              data = data)
```

```
summary(modell1)
```

```
##                                Df Sum Sq Mean Sq F value
## as.factor(chocolate_level)    4    42    10.4    0.249
## as.factor(time_level)         2   326   163.2    3.912
## as.factor(age_range)          4  1693   423.2   10.145
## as.factor(chocolate_level):as.factor(time_level)  8    243    30.4    0.730
## Residuals                    131  5464    41.7
##                                Pr(>F)
## as.factor(chocolate_level)    0.9095
## as.factor(time_level)         0.0224 *
## as.factor(age_range)          3.47e-07 ***
## as.factor(chocolate_level):as.factor(time_level)  0.6651
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

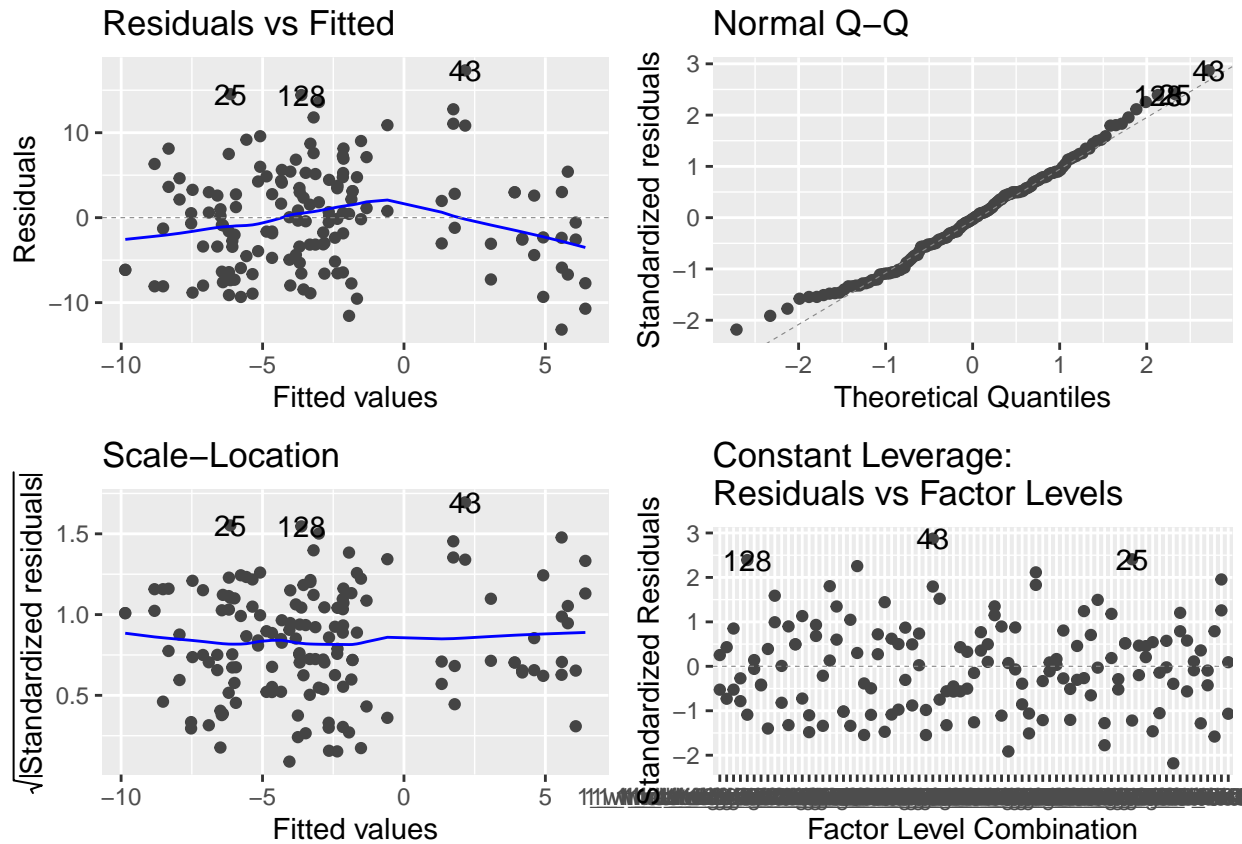


## Latex table for ANOVA Summary

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
chocolate_level	4.0000	41.6257	10.4064	0.2495	0.9095
time_level	2.0000	326.3177	163.1589	3.9118	0.0224
age_range	4.0000	1692.6137	423.1534	10.1452	3.47e-07
chocolate_level:time_level	8.0000	243.4743	30.4343	0.7297	0.6651
Residuals	131.0000	5463.9643	41.7097		

## Residual Plots

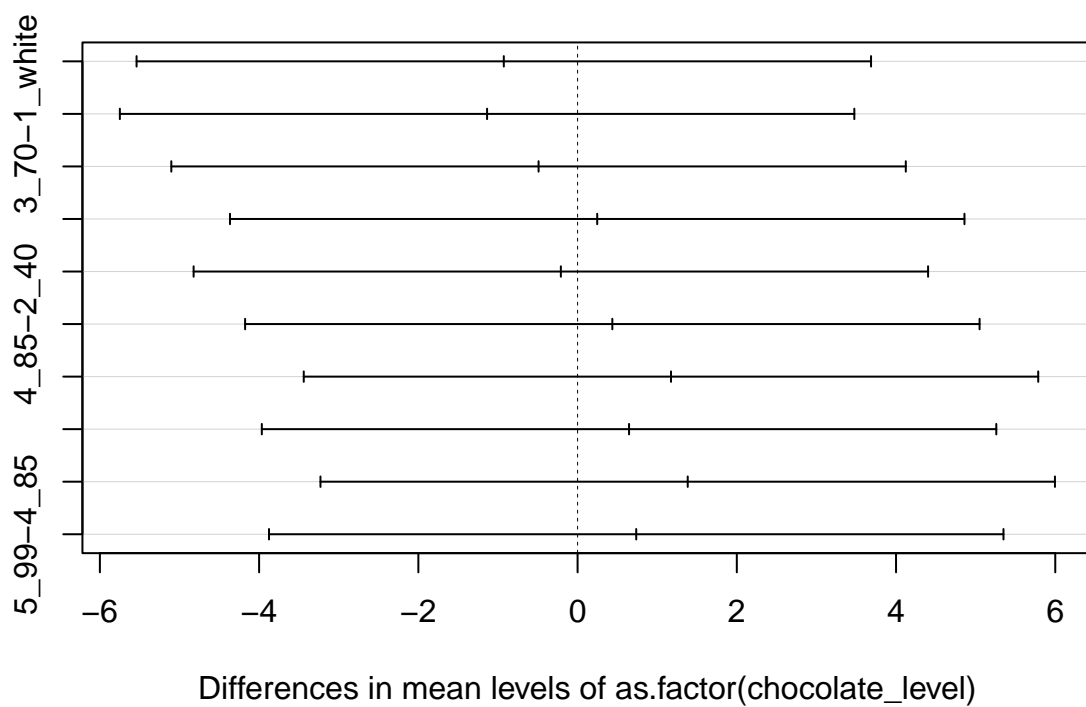
```
library(ggfortify)
par(mfrow = c(2,2))
autoplot(model1)
```



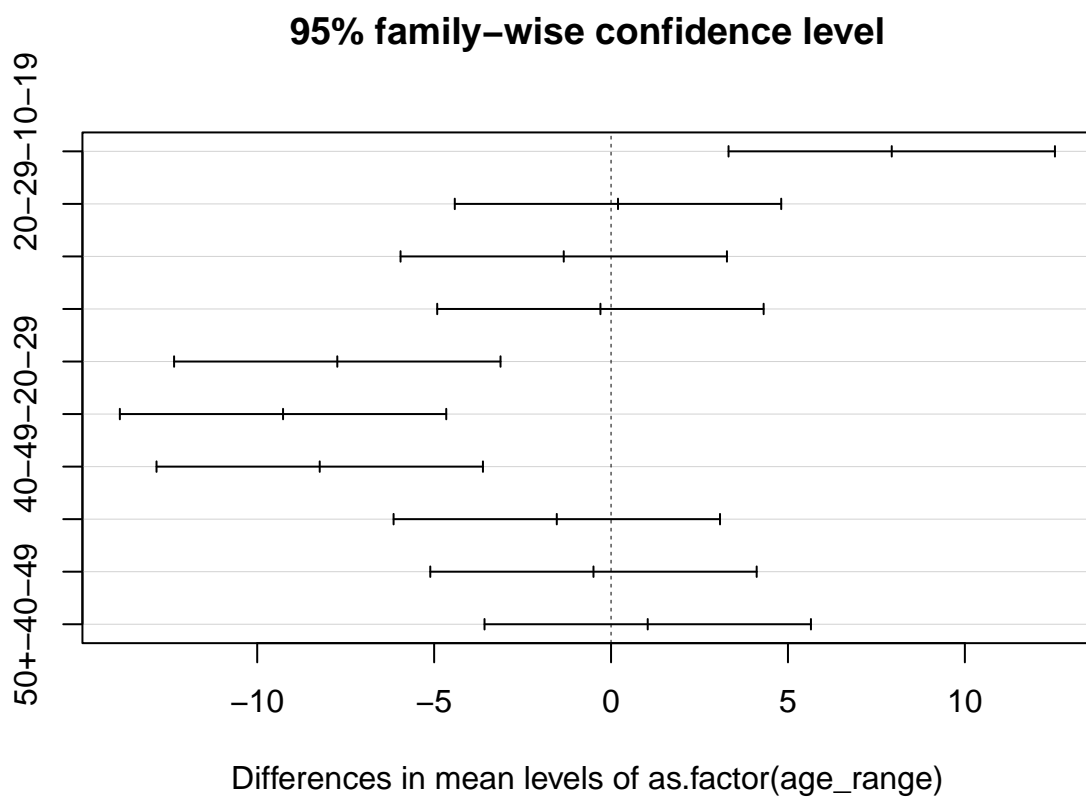
## Post Hoc Tukey

```
tukey <- TukeyHSD(model1)
plot(TukeyHSD(model1))
```

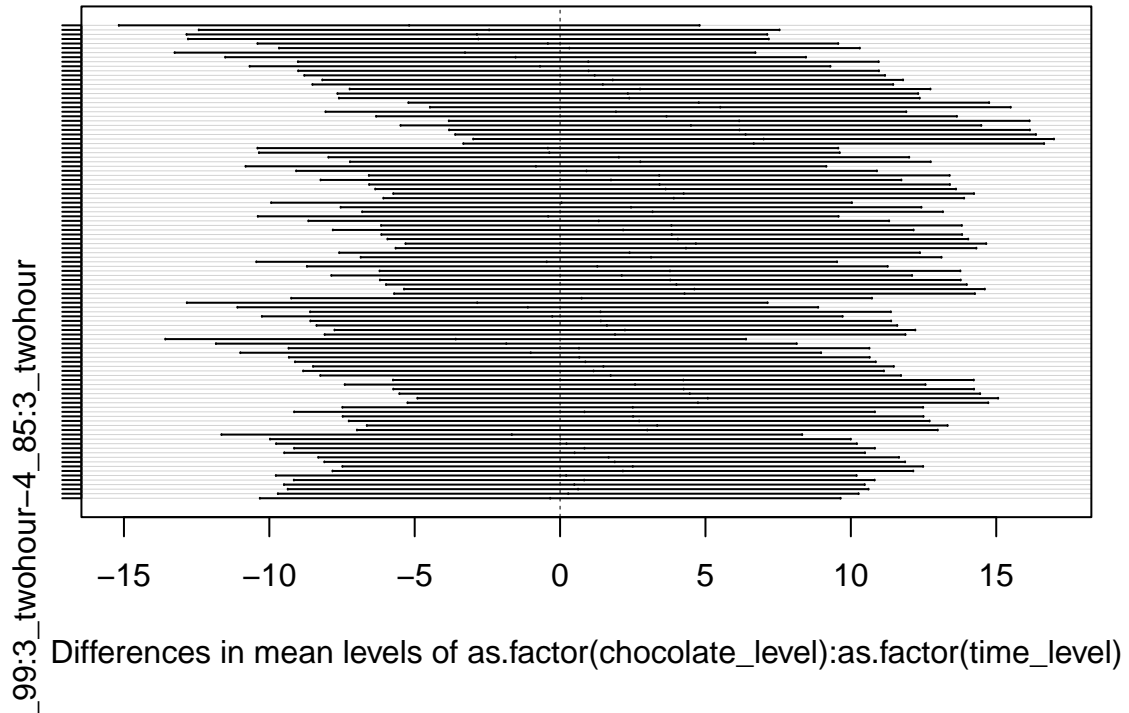
### 95% family-wise confidence level







## 95% family-wise confidence level



## Latex Tables for Post Hoc Tukey

Comparison	Difference	Lower	Upper	P Value Adjusted
right_after-one_hour	-1.874	-4.936077	1.188077	0.3180854
two_hours-one_hour	1.738	-1.324077	4.800077	0.3726983
two_hours-right_after	3.612	0.5499234	6.674077	<b>0.0162804</b>

Pairwise comparisons of time levels with adjusted p-values

Comparison	Difference	Lower	Upper	P Value Adjusted
20-29-10-19	7.9333333	3.320721	12.545946	<b>0.0000496</b>
30-39-10-19	0.1966667	-4.415946	4.809279	0.9999558
40-49-10-19	-1.3366667	-5.949279	3.275946	0.9297008
50+-10-19	-0.3000000	-4.912612	4.312612	0.9997627
30-39-20-29	-7.7366667	-12.349279	-3.124054	<b>0.0000810</b>
40-49-20-29	-9.2700000	-13.882612	-4.657388	<b>0.0000014</b>
50+-20-29	-8.2333333	-12.845946	-3.620721	<b>0.0000231</b>
40-49-30-39	-1.5333333	-6.145946	3.079279	0.8889256
50+-30-39	-0.4966667	-5.109279	4.115946	0.9982651
50+-40-49	1.0366667	-3.575946	5.649279	0.9713838

Pairwise comparisons of age ranges with adjusted p-values