

ASDS5301_FINAL_PROJECT

2023-12-06

TWO WAY ANOVA TEST

Team 5

Abhishek Goudar (1001916610)

JeyaSooriya Sarvanan (1002186838)

```
df_tooth_growth <- data.frame(ToothGrowth)

# Structure of the dataset
str(df_tooth_growth)

## 'data.frame': 60 obs. of 3 variables:
## $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

head(df_tooth_growth[c(28, 29, 30, 31, 32, 33), ],
      n = 6)

##      len supp dose
## 28 21.5   VC  2.0
## 29 23.3   VC  2.0
## 30 29.5   VC  2.0
## 31 15.2   OJ  0.5
## 32 21.5   OJ  0.5
## 33 17.6   OJ  0.5

##Dataset summary

summary(df_tooth_growth)

##      len      supp      dose
## Min.   : 4.20   OJ:30   Min.    :0.500
## 1st Qu.:13.07   VC:30   1st Qu.:0.500
## Median :19.25           Median :1.000
## Mean   :18.81           Mean   :1.167
## 3rd Qu.:25.27           3rd Qu.:2.000
## Max.   :33.90           Max.    :2.000

##Descriptive stats in conjunction to both factors
require(dplyr)

## Loading required package: dplyr

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

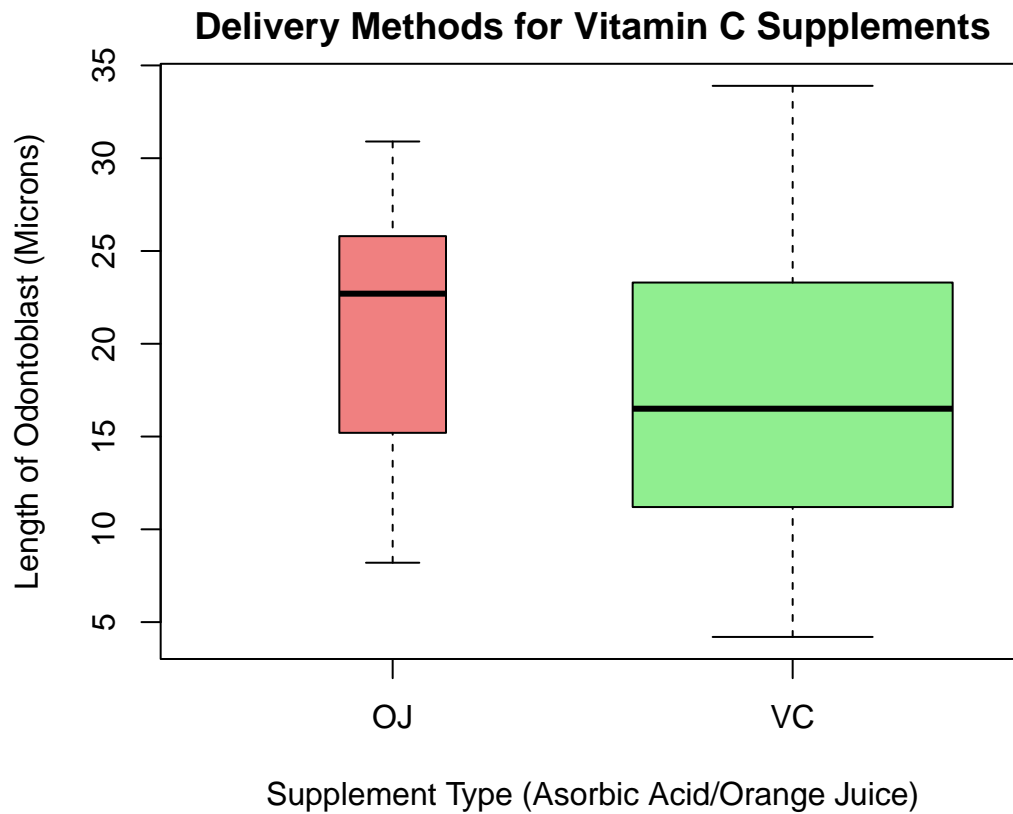
```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

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

```
library(dplyr)
group_by(df_tooth_growth, supp, dose) %>%
  summarise(
    count = n(),
    mean = mean(len, na.rm = TRUE),
    sd = sd(len, na.rm = TRUE),
    .groups = "drop"
  )
```

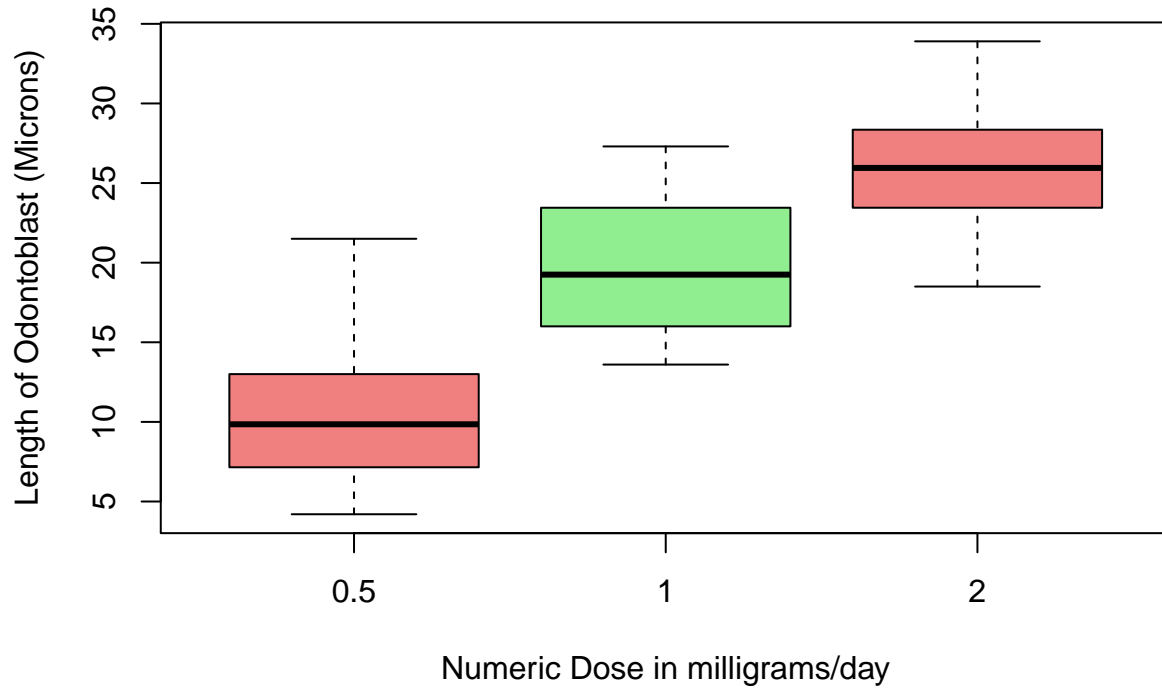
```
## # A tibble: 6 x 5
##   supp   dose count  mean    sd
##   <fct> <dbl> <int> <dbl> <dbl>
## 1 OJ     0.5    10 13.2  4.46
## 2 OJ     1      10 22.7  3.91
## 3 OJ     2      10 26.1  2.66
## 4 VC     0.5    10  7.98  2.75
## 5 VC     1      10 16.8  2.52
## 6 VC     2      10 26.1  4.80
```

```
## Boxplot for Delivery Methods of Supplement
par(mar = c(5, 5, 2, 5))
boxplot(len~supp, data=df_tooth_growth,
        notch=FALSE,
        col=c("lightcoral", "lightgreen")),
main="Delivery Methods for Vitamin C Supplements",
xlab="Supplement Type (Asorbic Acid/Orange Juice)",
ylab="Length of Odontoblast (Microns)",
width = c(0.5, 1.5)*2,
height = c(0.5, 1.5)
)
```



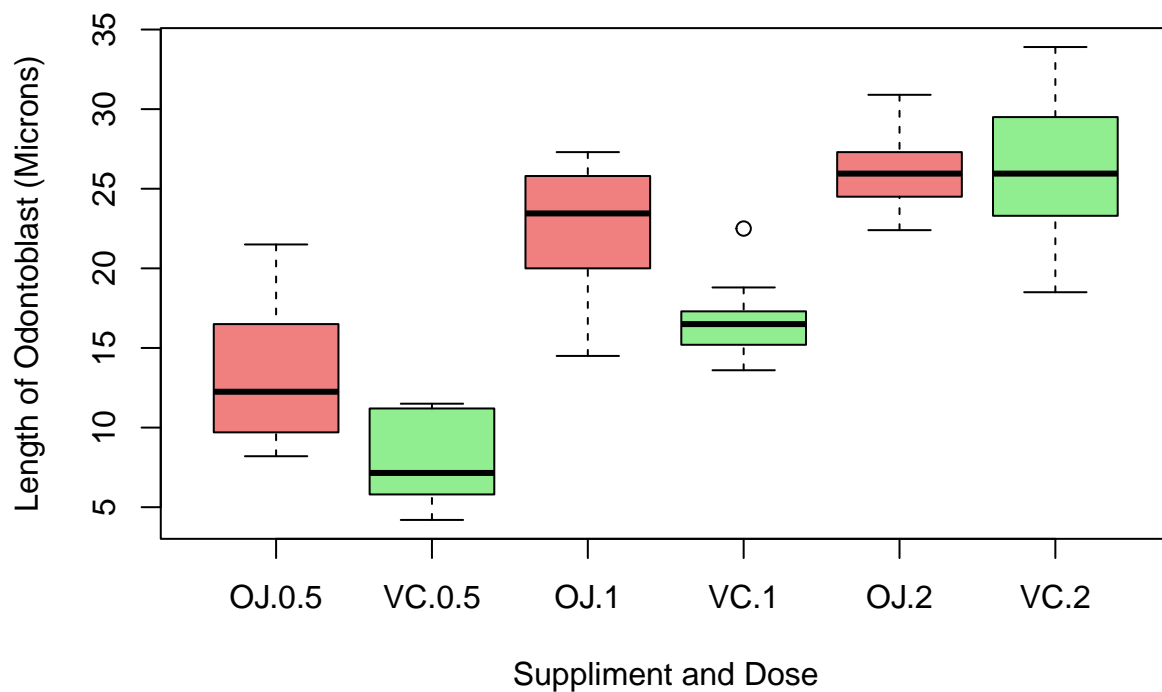
```
boxplot(len~dose, data=df_tooth_growth, notch=FALSE,  
        col=c("lightcoral","lightgreen"),  
        main="Three Dose Levels of Vitamin C (0.5, 1, and 2)",  
        xlab="Numeric Dose in milligrams/day",  
        ylab="Length of Odontoblast (Microns)"  
)
```

Three Dose Levels of Vitamin C (0.5, 1, and 2)



```
boxplot(len~supp*dose, data=df_tooth_growth, notch=FALSE,
        col=c("lightcoral","lightgreen")),
main="Odontoblast Growth",
xlab="Suppliment and Dose",
ylab="Length of Odontoblast (Microns)")
```

Odontoblast Growth



```
require(ggplot2)

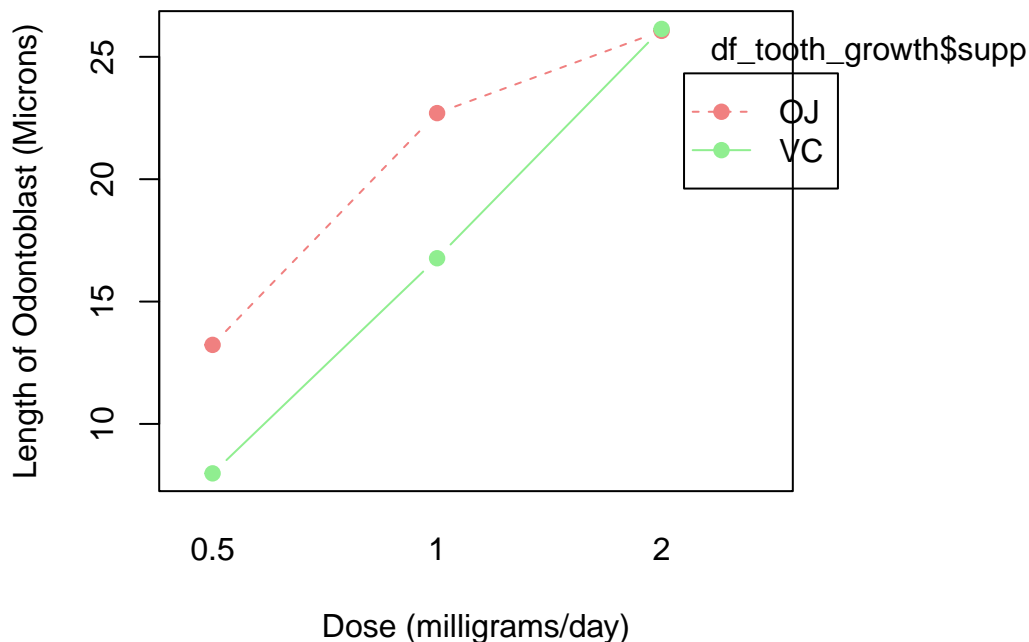
## Loading required package: ggplot2

require(dplyr)
library(ggplot2)
library(dplyr)

# Convert the 'supp' and 'dose' variables to factors
df_tooth_growth$supp <- as.factor(df_tooth_growth$supp)
df_tooth_growth$dose <- as.factor(df_tooth_growth$dose)

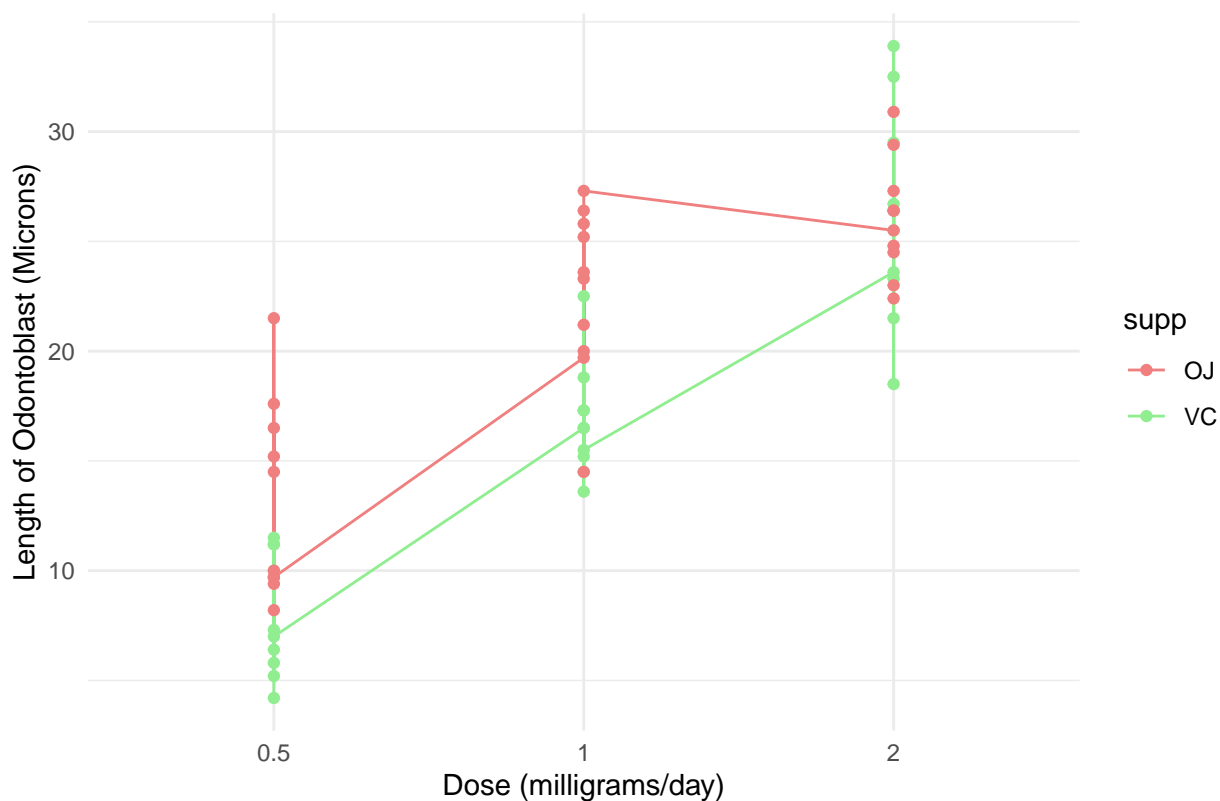
par(mar = c(8, 8, 2, 8))

# Interaction Plot
interaction.plot(df_tooth_growth$dose, df_tooth_growth$supp,
  df_tooth_growth$len,
  col = c("lightcoral", "lightgreen"),
  type = "b", fixed = TRUE,
  xlab = "Dose (milligrams/day)",
  ylab = "Length of Odontoblast (Microns)",
  pch = 19, leg.bty = "o")
```



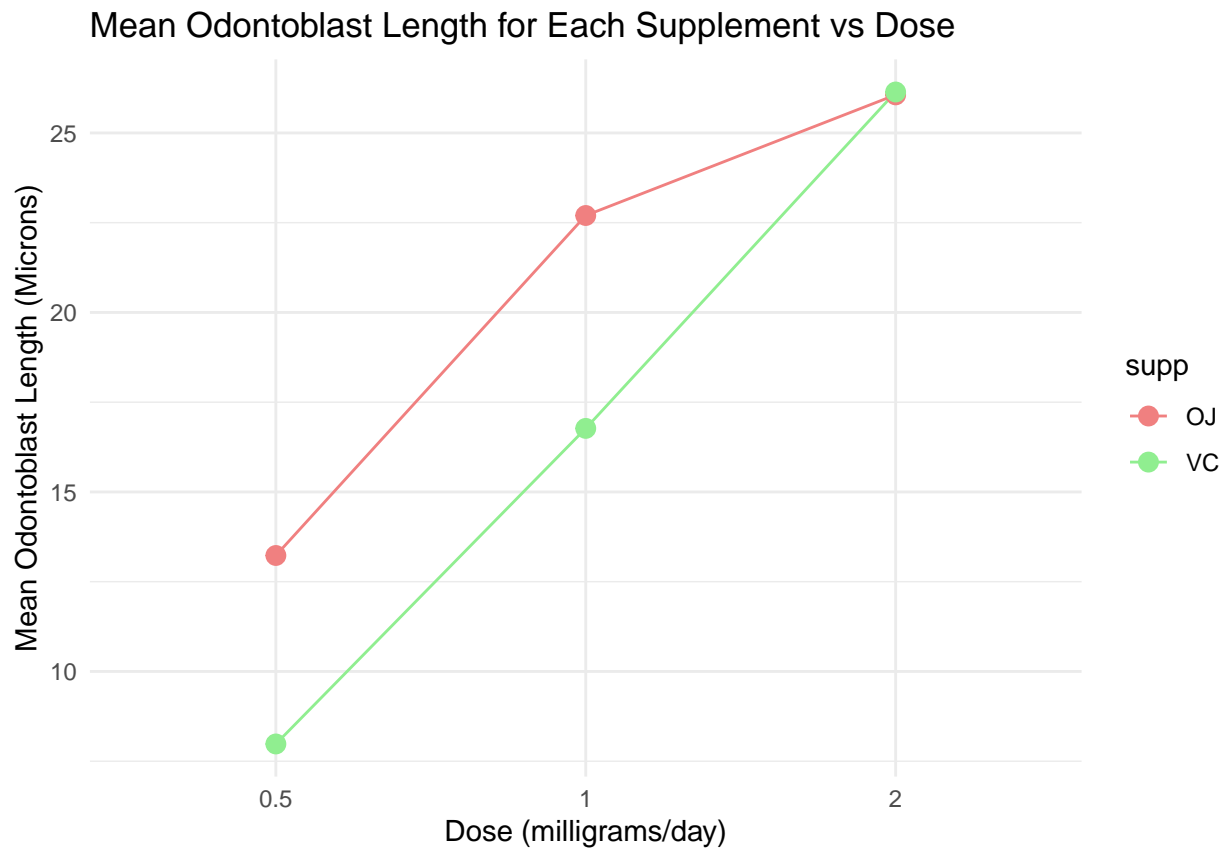
```
# Main Effects Plot for 'dose'
ggplot(df_tooth_growth, aes(x = dose, y = len, group = supp, color = supp)) +
  geom_line() +
  geom_point() +
  labs(title = "Main Effects Plot for Dose",
    x = "Dose (milligrams/day)", y = "Length of Odontoblast (Microns)") +
  scale_color_manual(values = c("lightcoral", "lightgreen"))
  ) + # Set custom colors
  theme_minimal()
```

Main Effects Plot for Dose

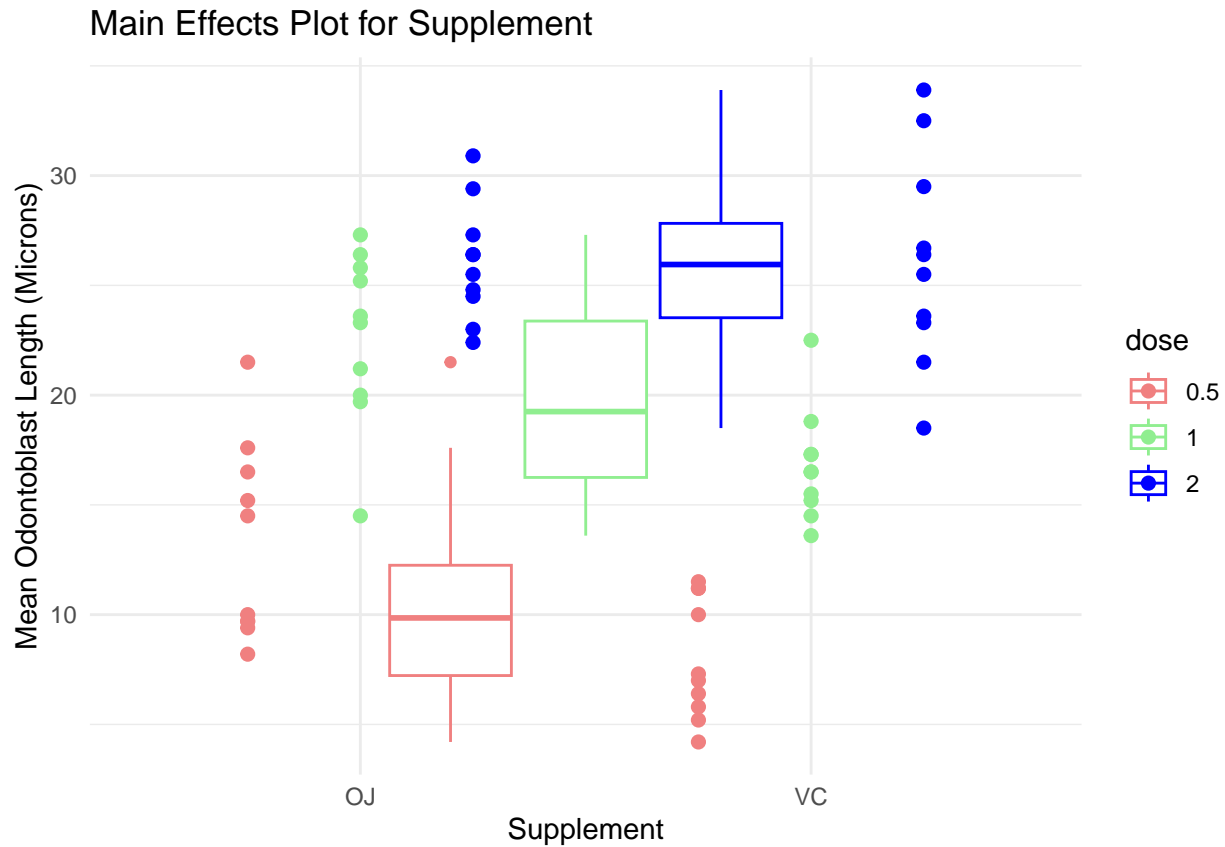


```
means_data <- df_tooth_growth %>%
  group_by(supp, dose) %>%
  summarize(mean_len = mean(len),
    .groups="drop")

# Line plot of means
ggplot(means_data, aes(x = dose, y = mean_len, group = supp, color = supp)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "Mean Odontoblast Length for Each Supplement vs Dose",
    x = "Dose (milligrams/day)", y = "Mean Odontoblast Length (Microns)") +
  scale_color_manual(values = c("lightcoral", "lightgreen")
    ) + # Set custom colors
  theme_minimal()
```



```
# Main Effects Plot for 'supp'
ggplot(df_tooth_growth, aes(x = supp, y = len, group = dose, color = dose)) +
  geom_boxplot() +
  geom_point(position = position_dodge(width = 0.75), size = 2) +
  labs(title = "Main Effects Plot for Supplement", x = "Supplement",
        y = "Mean Odontoblast Length (Microns)") +
  scale_color_manual(values = c("lightcoral", "lightgreen", "blue")
                    ) + # Set custom colors
  theme_minimal()
```



Null Hypothesis:

Alternate Hypothesis:

Test for Assumptions:

1. Shapiro-Wilk normality test to check for Normality:

Null Hypothesis (H0): The data in the sample is normally distributed.

Alternative Hypothesis (Ha): The data in the sample is not normally distributed.

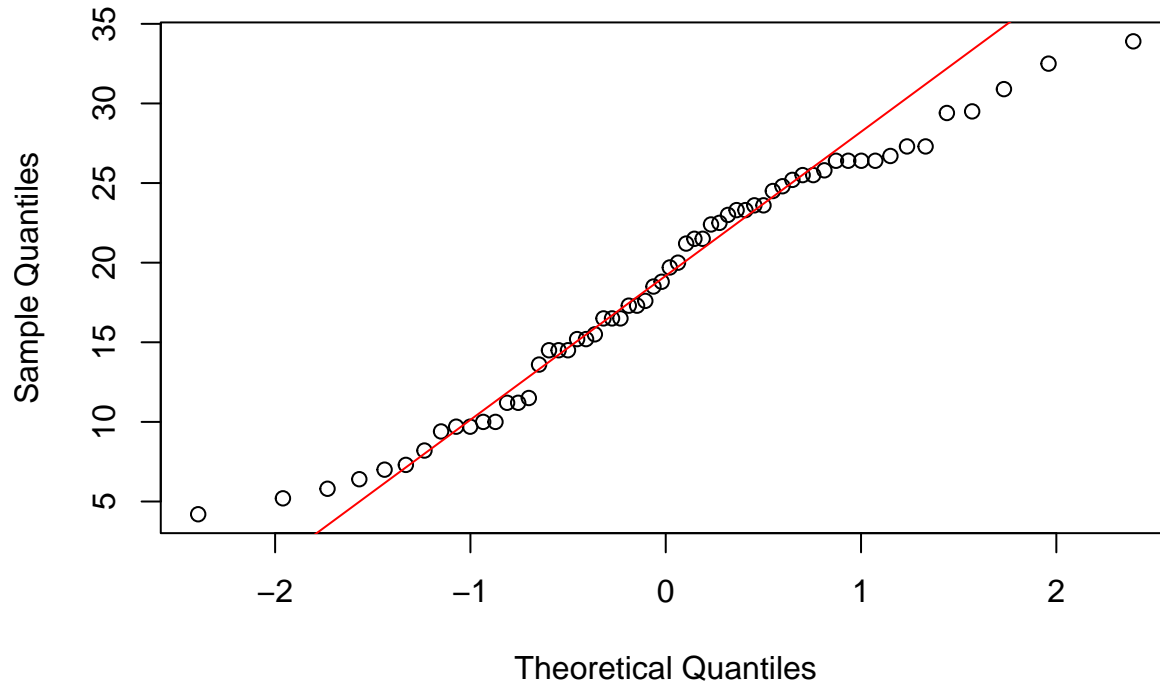
Used the below code to test for the Normality which is Shapiro.Test. Taken alpha = 0.05

```
shapiro.test(df_tooth_growth$len)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_tooth_growth$len
## W = 0.96743, p-value = 0.1091
```

```
qqnorm(df_tooth_growth$len)
qqline(df_tooth_growth$len, col = "red")
```


Normal Q-Q Plot



2.Ho-

mogeneity of variance Test

```
require(car)
```

```
## Loading required package: car
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##      recode
```

```
library(car)
# Levene test for Interaction Term
#levels(ToothGrowth$supp)
# Convert 'dose' to a factor with specific levels
df_tooth_growth$dose <- factor(df_tooth_growth$dose, levels = c("0.5", "1", "2"))

# Perform Levene test
leveneTest(len ~ supp * dose, data = df_tooth_growth)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  5  1.7086 0.1484
##      54
```

```
# Levene test for factor1
leveneTest(len ~ supp, data = df_tooth_growth)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
```

```
## group 1 1.2136 0.2752
##      58
```

```
# Levene test for factor2
```

```
leveneTest(len ~ dose, data = df_tooth_growth)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 2 0.6457 0.5281
```

```
##      57
```

3. Independence

```
# Create a contingency table
```

```
contingency_table <- table(df_tooth_growth$supp, df_tooth_growth$dose)
```

```
# Conduct Chi-squared test for independence
```

```
chi_squared_test <- chisq.test(contingency_table)
```

```
# Print the results
```

```
print(chi_squared_test)
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: contingency_table
```

```
## X-squared = 0, df = 2, p-value = 1
```

Since the Sample contains different categories within the corresponding features, we can confirm that the these features are independent of each other.

```
two_way_anova <- aov(len ~ supp * factor(dose), data = df_tooth_growth)
```

```
summary(two_way_anova)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## supp          1  205.4    205.4   15.572 0.000231 ***
## factor(dose)   2 2426.4   1213.2   92.000 < 2e-16 ***
## supp:factor(dose) 2  108.3     54.2    4.107 0.021860 *
## Residuals     54   712.1     13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(two_way_anova)
```

```
## Tukey multiple comparisons of means
```

```
## 95% family-wise confidence level
```

```
##
```

```
## Fit: aov(formula = len ~ supp * factor(dose), data = df_tooth_growth)
```

```
##
```

```
## $supp
```

```
##      diff      lwr      upr    p adj
## VC-0J -3.7 -5.579828 -1.820172 0.0002312
```

```
##
```

```
## $`factor(dose)`
```

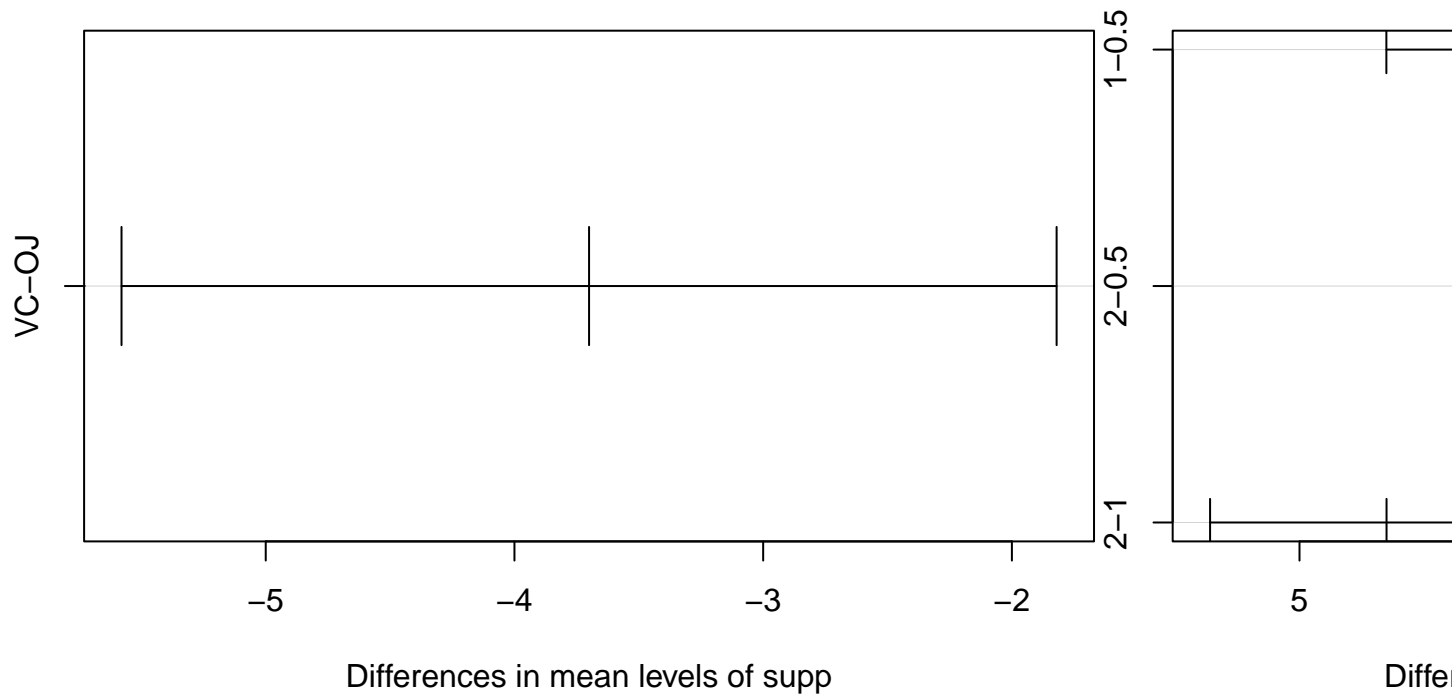
```
##      diff      lwr      upr    p adj
## 1-0.5  9.130  6.362488 11.897512 0.0e+00
## 2-0.5 15.495 12.727488 18.262512 0.0e+00
```

```
## 2-1      6.365  3.597488  9.132512  2.7e-06
##
## $`supp:factor(dose)`
##          diff          lwr          upr      p adj
## VC:0.5-OJ:0.5 -5.25 -10.048124 -0.4518762 0.0242521
## OJ:1-OJ:0.5   9.47   4.671876 14.2681238 0.0000046
## VC:1-OJ:0.5   3.54  -1.258124  8.3381238 0.2640208
## OJ:2-OJ:0.5  12.83   8.031876 17.6281238 0.0000000
## VC:2-OJ:0.5  12.91   8.111876 17.7081238 0.0000000
## OJ:1-VC:0.5  14.72   9.921876 19.5181238 0.0000000
## VC:1-VC:0.5   8.79   3.991876 13.5881238 0.0000210
## OJ:2-VC:0.5  18.08  13.281876 22.8781238 0.0000000
## VC:2-VC:0.5  18.16  13.361876 22.9581238 0.0000000
## VC:1-OJ:1    -5.93 -10.728124 -1.1318762 0.0073930
## OJ:2-OJ:1     3.36  -1.438124  8.1581238 0.3187361
## VC:2-OJ:1     3.44  -1.358124  8.2381238 0.2936430
## OJ:2-VC:1     9.29   4.491876 14.0881238 0.0000069
## VC:2-VC:1     9.37   4.571876 14.1681238 0.0000058
## VC:2-OJ:2     0.08  -4.718124  4.8781238 1.0000000
```

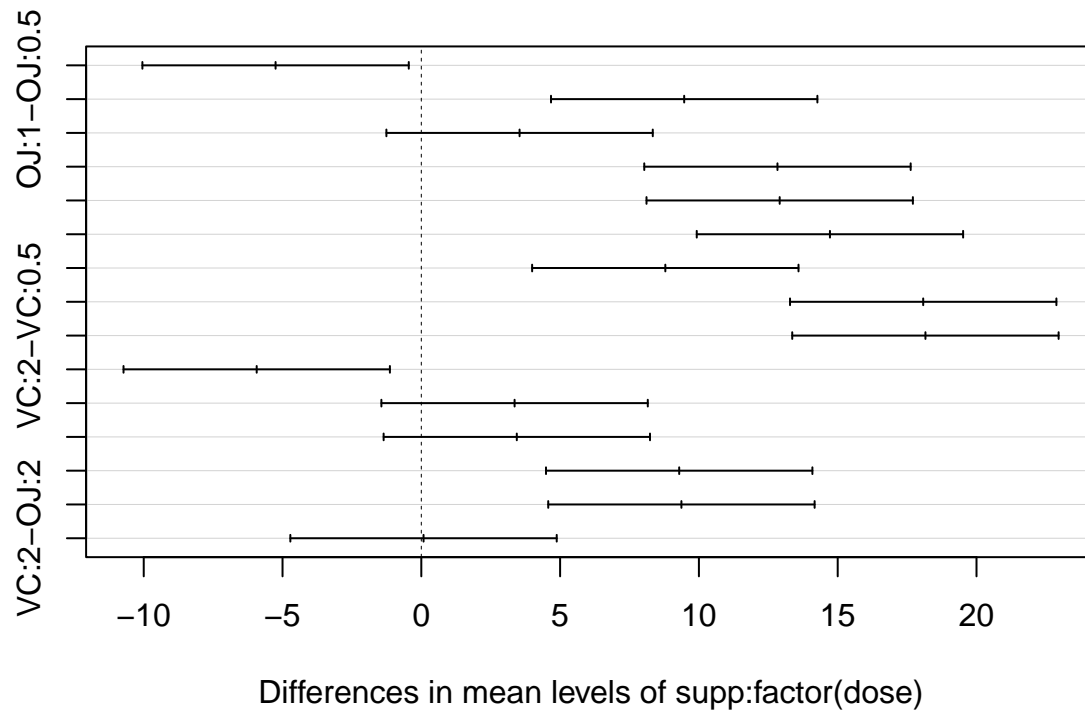
```
plot(TukeyHSD(two_way_anova))
```

95% family-wise confidence level

95%



95% family-wise confidence level



```
require(rstatix)
```

```
## Loading required package: rstatix
```

```
##
```

```
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
## filter
```

```
library(rstatix)
```

```
eta_squared(two_way_anova)
```

```
##          supp      factor(dose) supp:factor(dose)
```

```
##    0.05948365    0.70286419    0.03137672
```

```
partial_eta_squared(two_way_anova)
```

```
##          supp      factor(dose) supp:factor(dose)
```

```
##    0.2238254    0.7731092    0.1320279
```