

Lab Assignment

Sarah Rafiq Shaikh

1/24/2022

1. Selecting and Viewing a dataset

```
#Viewing all the built-in data sets in R and Selecting one of them
data()

#Loading , My data set is a small dataset which is about The Effect of Vitamin C on Tooth Growth in Guinea Pigs
data("ToothGrowth")

# names of the variables used in the data set that is len , supp and dose
names(ToothGrowth)

## [1] "len" "supp" "dose"
```

2. Exploring the Selected dataset

```
#printing top few values of the selected data set (only 6 values)
head(ToothGrowth)

##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5

#printing last 10 rows of selected data set
tail(ToothGrowth, n=10)

##      len supp dose
## 51 25.5   OJ   2
```

##	52	26.4	OJ	2
##	53	22.4	OJ	2
##	54	24.5	OJ	2
##	55	24.8	OJ	2
##	56	30.9	OJ	2
##	57	26.4	OJ	2
##	58	27.3	OJ	2
##	59	29.4	OJ	2
##	60	23.0	OJ	2

```
#This is a data frame with 60 observations on 3 variables meaning 60 rows (to  
oths) 3 cols (variables) (Printing Both rows and cols)
```

```
dim(ToothGrowth)
```

```
## [1] 60 3
```

```
# Printing Dimensions another time
```

```
print(dim(ToothGrowth))
```

```
## [1] 60 3
```

#60 teeth(obs) (Printing Only row) of the selected data set

```
nrow(ToothGrowth)
```

```
## [1] 60
```

#3 variables (Printing Only cols) of the selected data set

```
ncol(ToothGrowth)
```

```
## [1] 3
```

#Categorical Data (variables) of the selected data set

ToothGrowth\$supp

[illegible]

```
## [26] VC VC VC VC VC OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ
OJ OJ
```

```
## [51] 0J 0J 0J 0J 0J 0J 0J 0J 0J 0J 0J
```

```
## Levels: OJ VC
```

#Converting categorical data to numeric data of the selected data set

```
as.numeric(ToothGrowth$supp)
```

[illegible]

```
## [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

3. Data Summarization

```
#Summarizing structure of data set
#The selected data set has 60 observations and 3 types of variables in the data set
1. len (numeric) - Tooth length
2. supp (factor) - Supplement type (VC or OJ)
3. dose (numeric) - Dose in milligrams
str(ToothGrowth)

## '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 ...

# Summarizing
#that tooth length is dependent variable, we have for the independent variables, supplement type and dosage
The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).
table(ToothGrowth$supp, ToothGrowth$dose)

##
##      0.5  1  2
##  OJ   10 10 10
##  VC   10 10 10

#Summary for all the variables of the data set
#The summary confirms that the supp and dose can be treated as factors. We can see that len ranges from 4.20 to 33.90.
summary(ToothGrowth)

##      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
```

4. Data Visualization

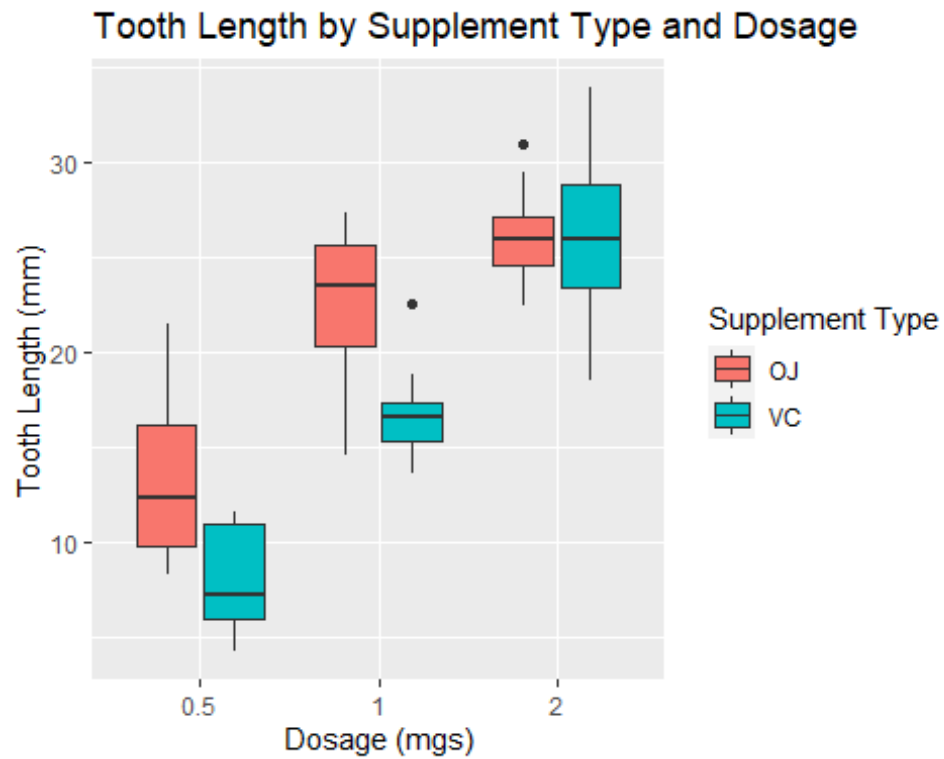
```
attach(ToothGrowth)
#install.packages("ggplot2") ## Loading required package: ggplot2

#Doing box plot of tooth length broken down by supplement type and dosage'
require(ggplot2)

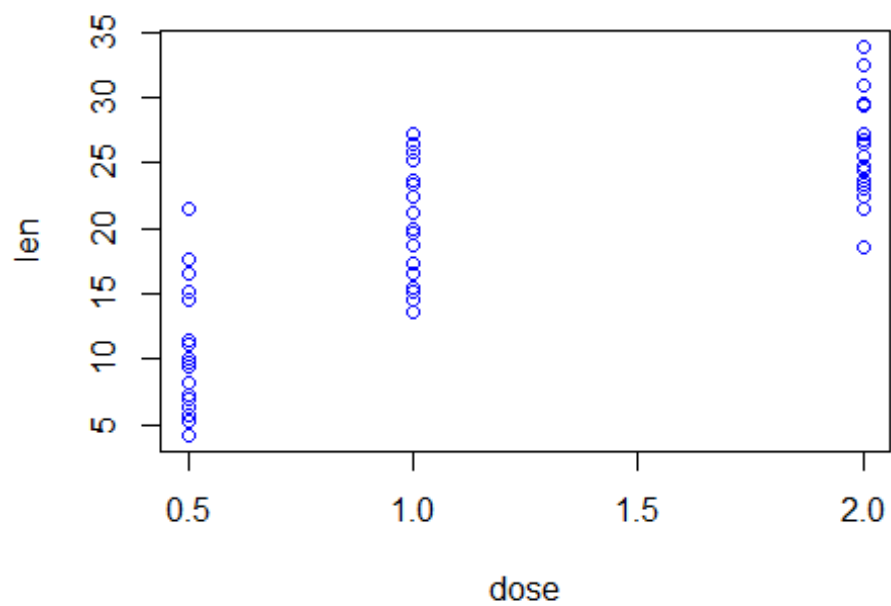
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.1.2

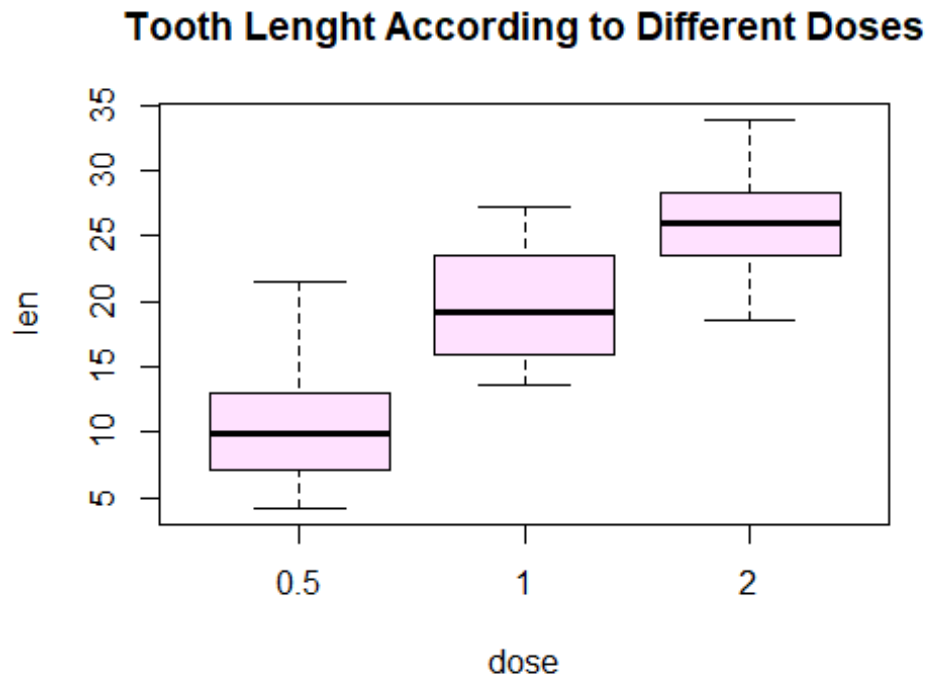
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=supp)) +
  geom_boxplot() +
  ggtitle('Tooth Length by Supplement Type and Dosage') +
  xlab('Dosage (mgs)') +
  ylab('Tooth Length (mm)') +
  guides(fill=guide_legend(title='Supplement Type'))
```



#Doing Scatter plot tooth Length broken by dose
`plot(len ~ dose , col="Blue")`

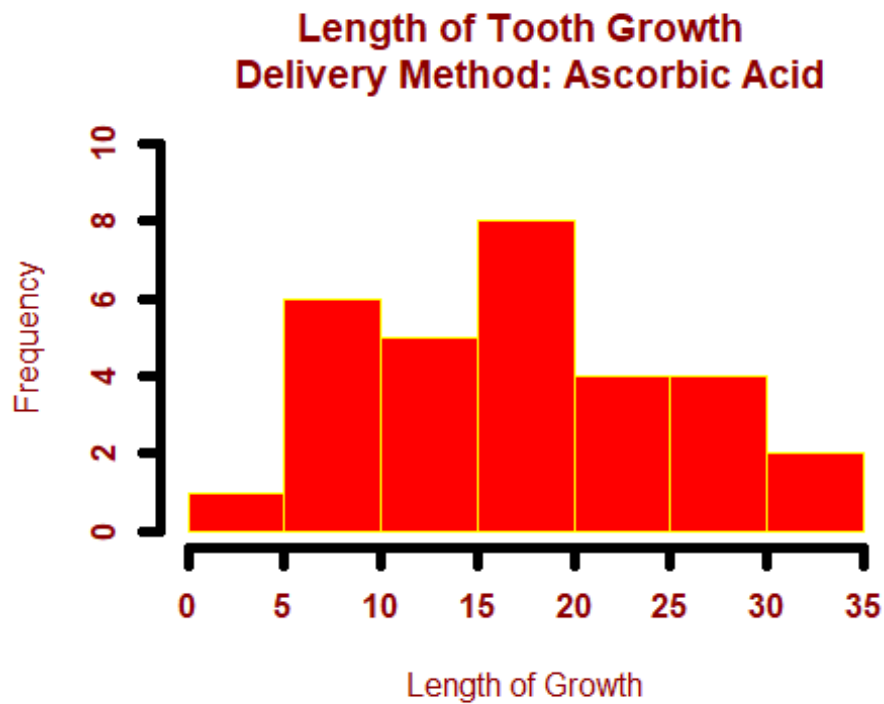


```
#Doing Simple boxplot of tooth Length broken down by supplement type and dose
boxplot(len ~ dose , main="Tooth Length According to Different Doses", col =
"thistle1")
```

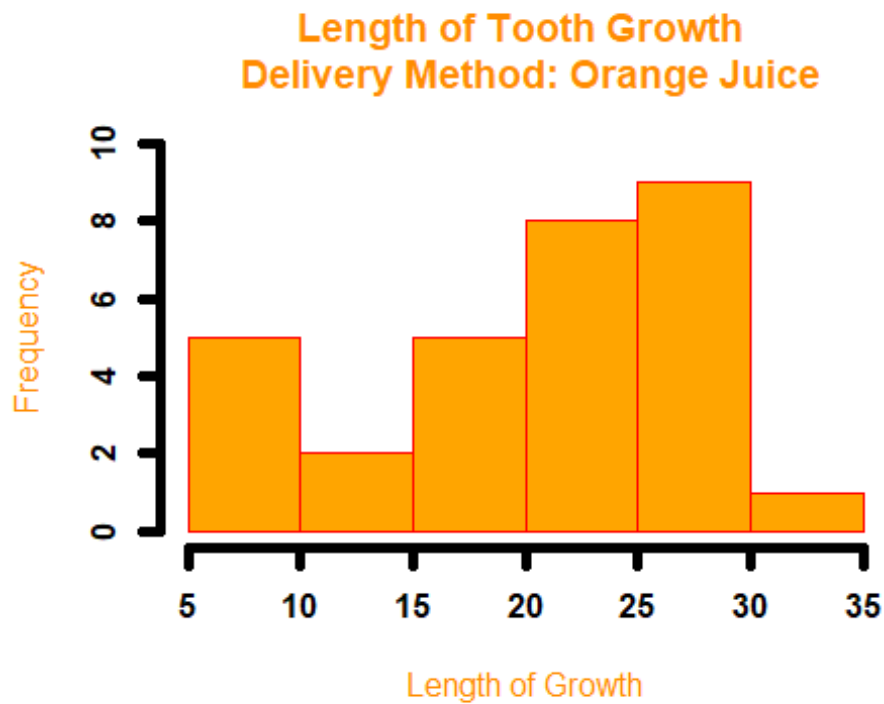


```
#separate the data set by delivery method
vctooth <- ToothGrowth[1:30, c(3,1)]
ojtooth <- ToothGrowth[31:60, c(3,1)]

#histograms of Length for the two delivery methods:
hist(vctooth$len,
     breaks="Sturges",
     border="yellow",
     font=2,
     lwd=5,
     cex=1.5,
     ylim=c(0,10),
     main = 'Length of Tooth Growth \n Delivery Method: Ascorbic Acid',
     xlab="Length of Growth",
     ylab="Frequency",
     col="red",
     col.main="darkred",
     col.lab="darkred",
     col.axis="darkred")
```

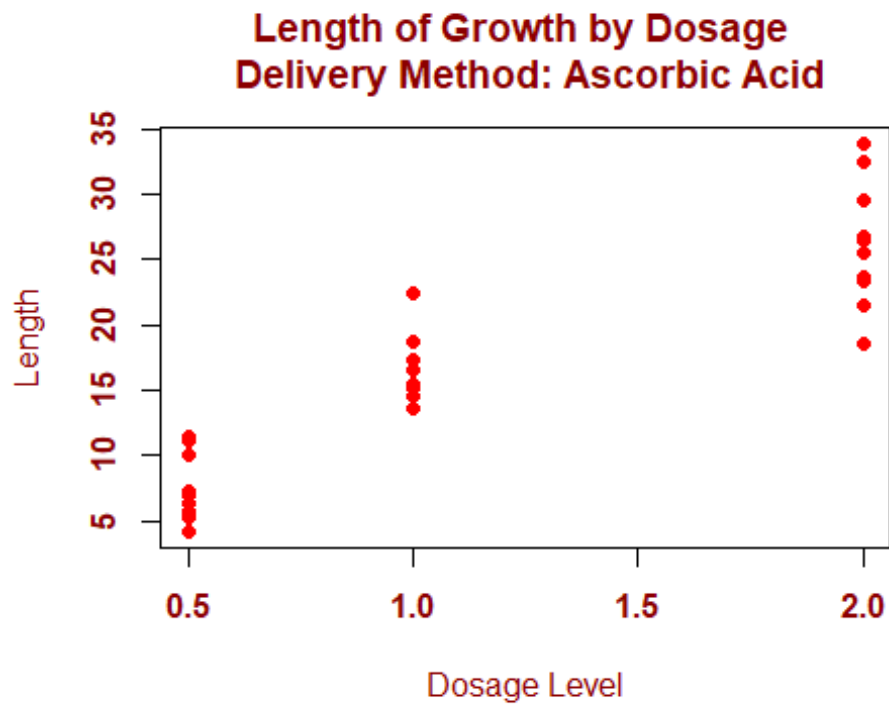


```
hist(ojtooth$len,  
     breaks="Sturges",  
     border="red",  
     font=2,  
     lwd=5,  
     cex=1.5,  
     ylim=c(0,10),  
     main = 'Length of Tooth Growth \n Delivery Method: Orange Juice',  
     xlab="Length of Growth",  
     ylab="Frequency",  
     col="orange",  
     col.main="darkorange",  
     col.lab="darkorange")
```

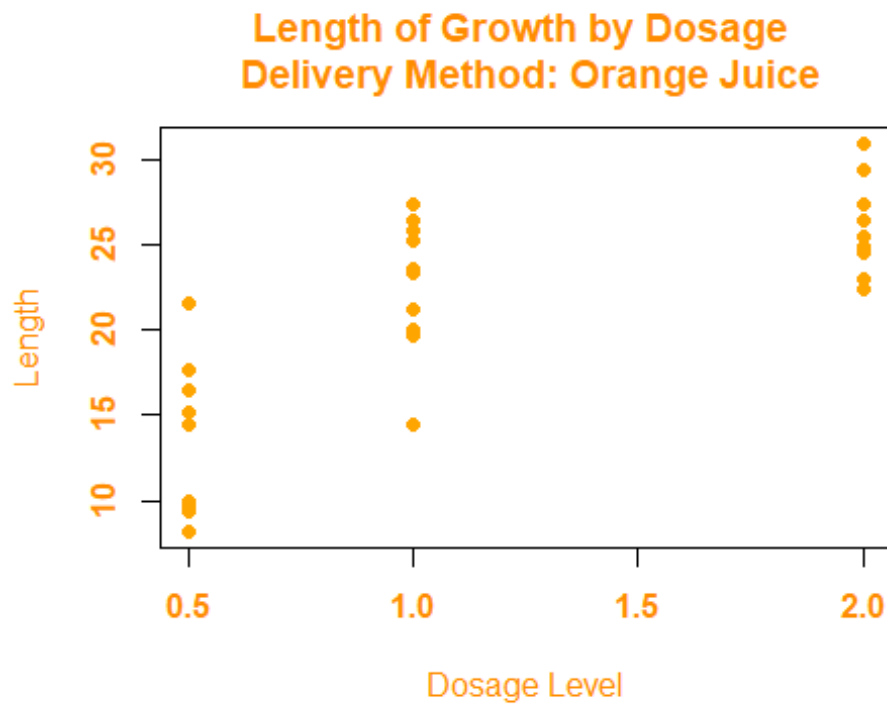


#breakdowns of Length by dose for each delivery method

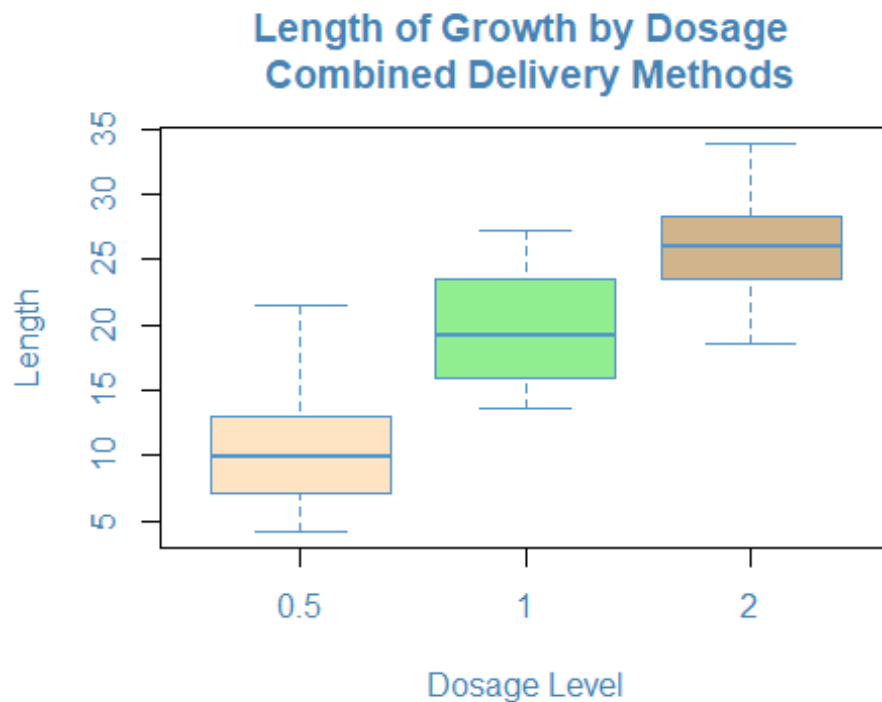
```
plot(vctooth,  
     main = 'Length of Growth by Dosage \n Delivery Method: Ascorbic Acid',  
     xlab="Dosage Level",  
     ylab="Length",  
     lwd=4,  
     pch=16,  
     font=2,  
     col="red",  
     col.main="darkred",  
     col.lab="darkred",  
     col.axis="darkred")
```

```
plot(ojtooth,
     main = 'Length of Growth by Dosage \n Delivery Method: Orange Juice',
     xlab="Dosage Level",
     ylab="Length",
     lwd=5,
     pch=16,
     font=2,
     col="orange",
     col.main="darkorange",
     col.lab="darkorange",
     col.axis="darkorange")
```



```
#box plots for length by dosage level regardless of delivery method:
plot(len ~ as.factor(dose),
     col=c("bisque1", "lightgreen", "tan"),
     xlab="Dosage Level",
     border="steelblue3",
     lwd=0.8,
     ylab="Length",
     col.lab="steelblue",
     col.axis="steelblue",
     col.main="steelblue",
     main="Length of Growth by Dosage \n Combined Delivery Methods",
     data=ToothGrowth)
```



#ALL the Above Plots are showing that :

#For 2.0 mg dosage, there appears to be no difference between OJ and VC.

#For 1.0 mg dosage, OJ definitely appears to promote higher tooth growth than VC

#For 0.5 mg dosage, OJ appears to promote higher tooth growth than VC., dosage , as the "dosage increases the tooth growth increases". In the specific case of the VC, the tooth growth has a linear relationship with dosage. The higher dosage (2.0mg) has less improvement in tooth growth with the OJ supplement. However, the OJ supplement generally induces more tooth growth than VC except at higher dosage (2.0 mg).

5. Correlation

```
# install.packages("corrplot") installing required package

sapply(ToothGrowth, is.numeric)

##   len  supp  dose
## TRUE FALSE  TRUE

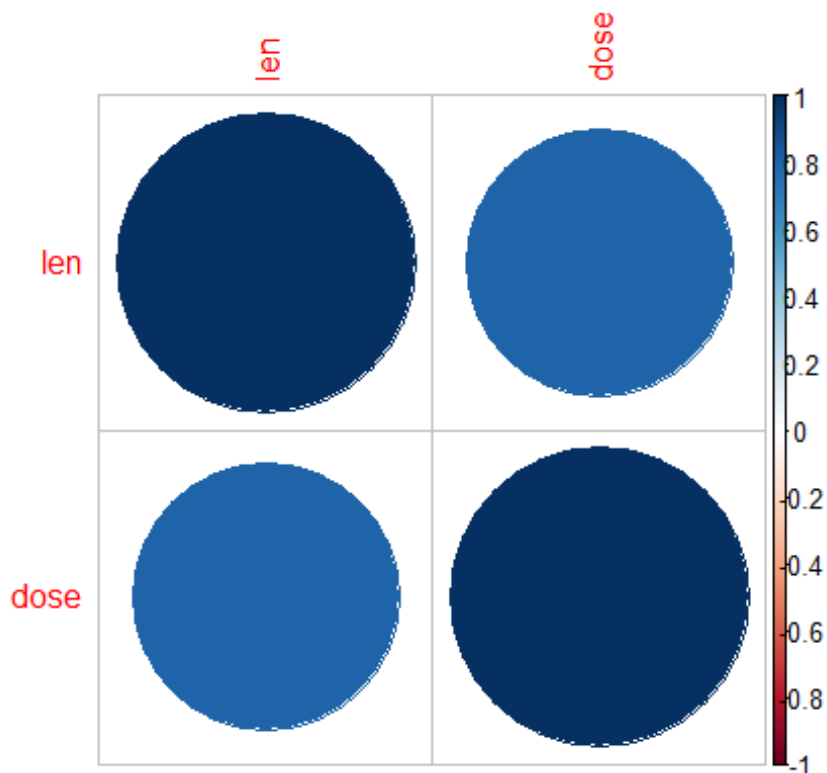
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.1.2
## corrplot 0.92 loaded

  cor(ToothGrowth[, unlist(lapply(ToothGrowth, is.numeric))])

##           len      dose
## len  1.0000000 0.8026913
## dose 0.8026913 1.0000000

  cor.mat.ToothGrowth = cor(ToothGrowth[, unlist(lapply(ToothGrowth, is.numer
ic))])
corrplot(cor.mat.ToothGrowth)
```



```

#From the plot we have concluded that:
#There obvious correlation between dose and Length of the teeth.
#"Larger dose results in larger tooth length, high and positive correlation
# there seems to be a positive correlation between dose and length. 2- Orange
juice seems more effective than Ascorbic Acid with Low (0.5) and medium (1.0)
doses. 3- With high doses (2.0), the median length is identical with OJ and V
C. 4- There seems to be a decreasing effectivity of Orange Juice: if we check
the median length, the increment of length from 0.5 to 1.0 is greater than th
e increment from 1.0 to 2.0). This attenuation effect is not observed in the
case of the ascorbic acid.

```

6. Confidence Interval

```

# install.packages("Rmisc") installing required package
library(Rmisc)

## Warning: package 'Rmisc' was built under R version 4.1.2

## Loading required package: lattice

## Loading required package: plyr

## Warning: package 'plyr' was built under R version 4.1.2

CI(ToothGrowth$dose, ci = 0.95)

##      upper      mean      lower
## 1.329122 1.166667 1.004212

# mean value lie in between the the upper and lower class interval which indi
cates accept the null hypothesis and mean is not significantly different fro
m 1.16, not different from 0

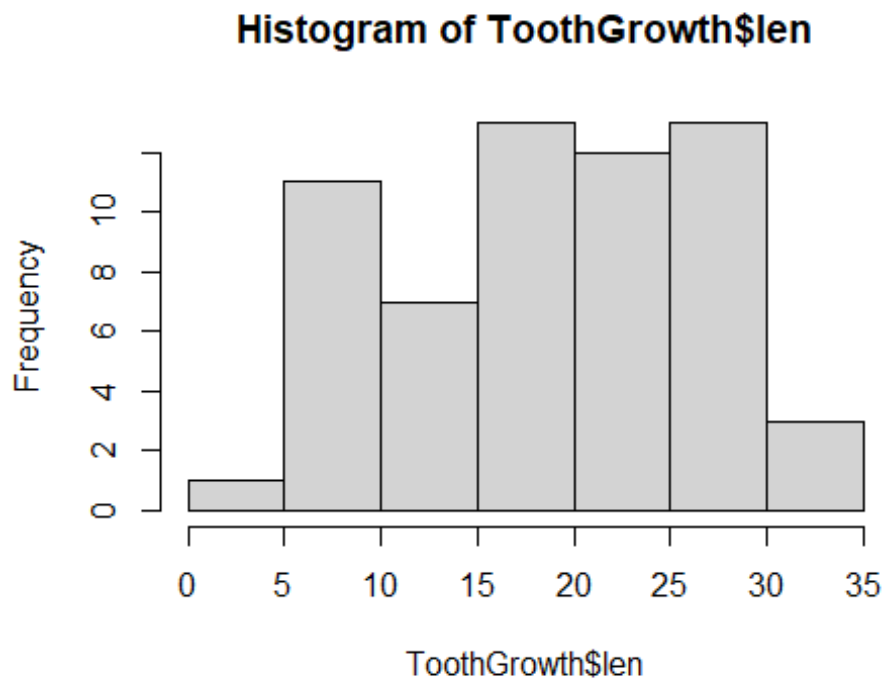
CI(ToothGrowth$len , ci = 0.95)

##      upper      mean      lower
## 20.78936 18.81333 16.83731

# mean value lie in between the the upper and lower class interval which indi
cates accept the null hypothesis and mean is not significantly different fro
m 18.81, not different from 0

##Visualization
#The chart below shows the distribution of tooth lengths. It seems to be some
what normally distributed
hist(ToothGrowth$len)

```



7. Hypothesis Testing

```
#install.packages("stats19") installing required package
library(stats19)

## Warning: package 'stats19' was built under R version 4.1.2

## Data provided under OGL v3.0. Cite the source and link to:
## www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

# One sample t-test

# to test: Is the mean value of dose differ from 1.167 or not?
t.test(mtcars$mpg, mu = 1.167 )

##
## One Sample t-test
##
## data:  mtcars$mpg
## t = 17.762, df = 31, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 1.167
## 95 percent confidence interval:
##  17.91768 22.26357
```

```
## sample estimates:
```

```
## mean of x
```

```
## 20.09062
```

```
t.test(mtcars$mpg, mu = 1.167, alternative = 'greater')
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: mtcars$mpg
```

```
## t = 17.762, df = 31, p-value < 2.2e-16
```

```
## alternative hypothesis: true mean is greater than 1.167
```

```
## 95 percent confidence interval:
```

```
## 18.28418 Inf
```

```
## sample estimates:
```

```
## mean of x
```

```
## 20.09062
```

#answer p value is less than 0.05 so null hypothesis is rejected

#In this case the 95% confidence intervals is zero, therefore we the null hypothesis: the mean values of the lengths per doses are not equal.

Two sample t-test for two independent samples

```
t.test(ToothGrowth$len, mu = 18.81)
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: ToothGrowth$len
```

```
## t = 0.0033755, df = 59, p-value = 0.9973
```

```
## alternative hypothesis: true mean is not equal to 18.81
```

```
## 95 percent confidence interval:
```

```
## 16.83731 20.78936
```

```
## sample estimates:
```

```
## mean of x
```

```
## 18.81333
```

#answer:First this is two tail test after this the mean value of weights differs from 18.81, and p value is more than 0.05 so null hypothesis is accepted

#In this case the 95% confidence intervals is zero, therefore we fail to reject the null hypothesis: the mean values of the lengths per doses are equal.

```
t.test(ToothGrowth$len, ToothGrowth$dose)
```

```
##
```

```
## Welch Two Sample t-test
```

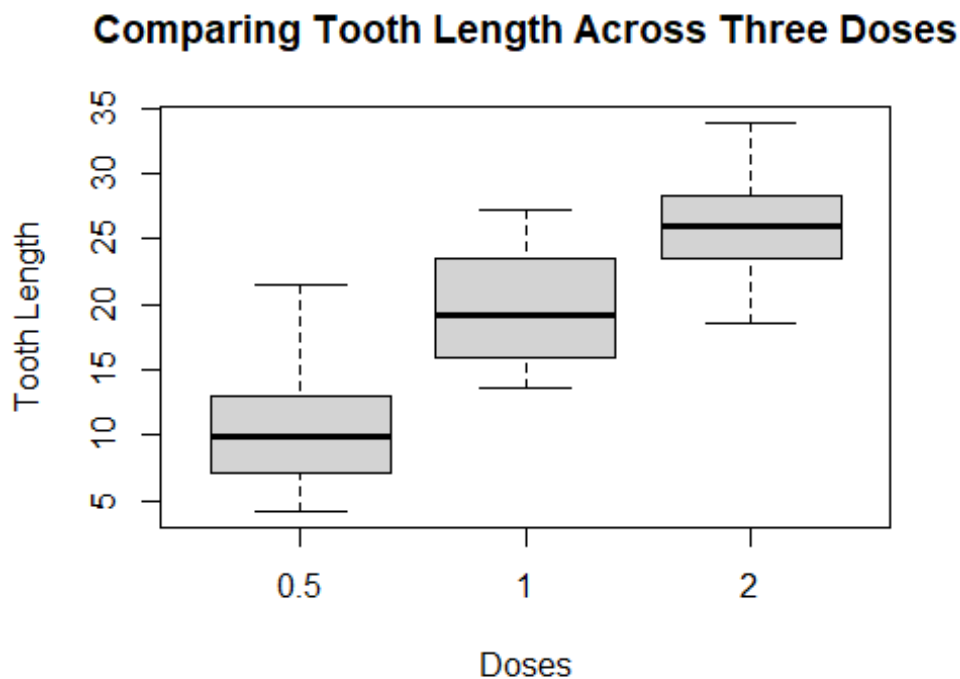
```
##
```

```
## data: ToothGrowth$len and ToothGrowth$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

#significantly different , both sides are positive and also p value is less than 0.05, alternative hypothesis is true

#Visualization

```
boxplot(ToothGrowth$len ~ ToothGrowth$dose, main="Comparing Tooth Length Across Three Doses", xlab="Doses", ylab="Tooth Length")
```



```
t.test(len ~ as.numeric(supp), data=ToothGrowth)

##
## Welch Two Sample t-test
##
## data: len by as.numeric(supp)
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group 1 and group
## 2 is not equal to 0
## 95 percent confidence interval:
## -0.1710156  7.5710156
```

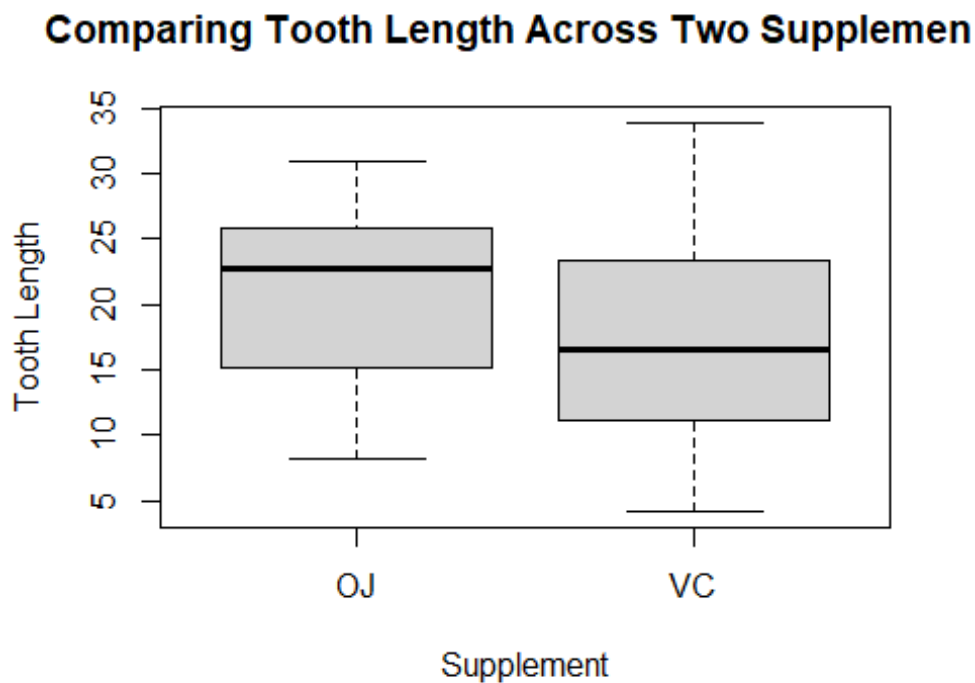


```
## sample estimates:  
## mean in group 1 mean in group 2  
##          20.66333          16.96333
```

The boxplot shows that the median length of teeth for the OJ delivery method is slightly higher than that for the VC method the p-value of the T-test is 0.06, which means this is higher than the 5% probability accepted for the 95% confidence interval. In addition, the confidence interval of the difference between the two means is -0.1710156 to 7.5710156, which (slightly) touches the zero. Which indicate that we fail to accept the null hypothesis

#Visualization

```
boxplot(ToothGrowth$len ~ ToothGrowth$supp, main="Comparing Tooth Length Across Two Supplements", xlab="Supplement", ylab="Tooth Length")
```



8. Chi - square Test

#apply the Chi-Square test to see test of association/ independence.

```
chisq.test(ToothGrowth$len,ToothGrowth$dose)
```

```
## Warning in chisq.test(ToothGrowth$len, ToothGrowth$dose): Chi-squared  
## approximation may be incorrect
```

```
##  
## Pearson's Chi-squared test  
##  
## data: ToothGrowth$len and ToothGrowth$dose  
## X-squared = 92.5, df = 84, p-value = 0.2464
```

#we have x-squared = 92.5 ,Since we get a p-Value greater than the significance level of 0.05, we accept the null hypothesis and conclude that the two variables are in fact independent and not associated with each other

#Creates a new variable Treat with 6 levels

```
ToothGrowth$Treat=with(ToothGrowth,interaction(supp,dose))
```

```
chisq.test(ToothGrowth$len , ToothGrowth$Treat)
```

```
## Warning in chisq.test(ToothGrowth$len, ToothGrowth$Treat): Chi-squared  
## approximation may be incorrect
```

```
##  
## Pearson's Chi-squared test  
##  
## data: ToothGrowth$len and ToothGrowth$Treat  
## X-squared = 223, df = 210, p-value = 0.2565
```

#we have x-squared = 223 ,Since we get a p-Value greater than the significance level of 0.05, we accept the null hypothesis and conclude that the two variables are in fact independent and not associated with each other

```
chisq.test(ToothGrowth$len,ToothGrowth$supp)
```

```
## Warning in chisq.test(ToothGrowth$len, ToothGrowth$supp): Chi-squared  
## approximation may be incorrect
```

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

```
## data: ToothGrowth$len and ToothGrowth$supp
## X-squared = 39.667, df = 42, p-value = 0.5739
```

#we have $x\text{-squared} = 39.667$,Since we get a $p\text{-Value}$ greater than the significance level of 0.05 , we accept the null hypothesis and conclude that the two variables are in fact independent and not associated with each other

##Warning Reason

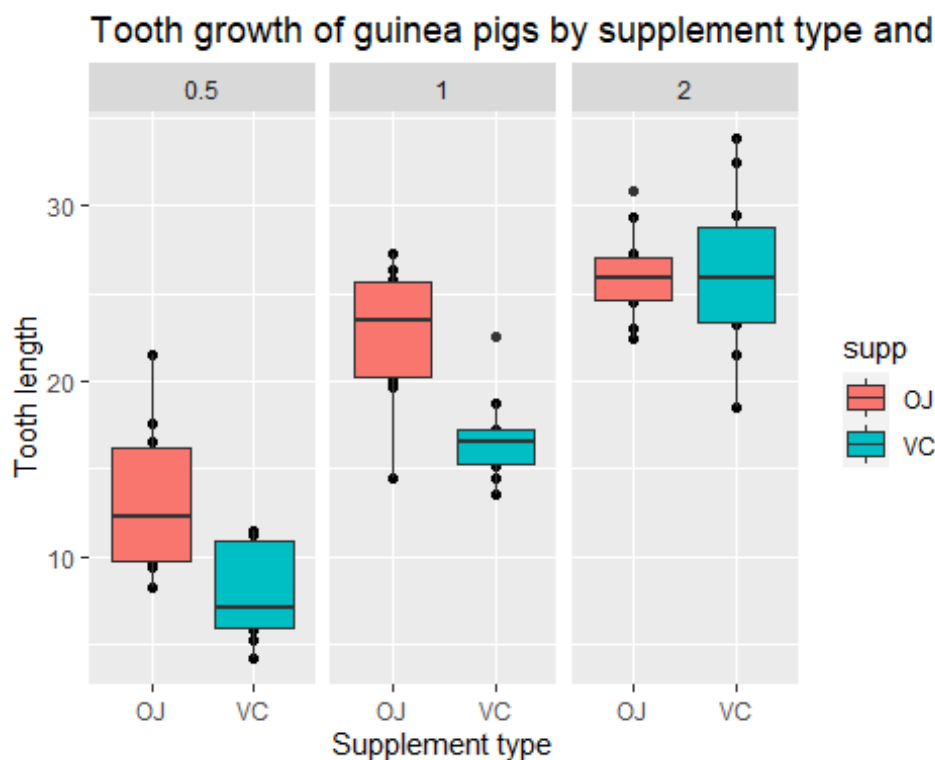
#It gave the warning because many of the expected values will be very small and therefore the approximations of p may not be right.

##Wrong p- values

Here the p approximation is incorrect because the length the teeth is highly dependent on both dose and supplements

##Showing the tooth length dependency on tooth dosage and supplement

```
qplot(supp,len,data=ToothGrowth, facets=~dose, main="Tooth growth of guinea pigs by supplement type and dosage (mg)",xlab="Supplement type", ylab="Tooth length") + geom_boxplot(aes(fill = supp))
```



#Length is highly dependent on supplement and dosage

#dosage, as the dosage increases the tooth growth increases. In the specific case of the VC, the tooth growth has a linear relationship with dosage. The higher dosage (2.0mg) has less improvement in tooth growth with the OJ supplement. However, the OJ supplement generally induces more tooth growth than VC except at higher dosage (2.0 mg).

9. Analysis of Variance

```
attach(ToothGrowth)

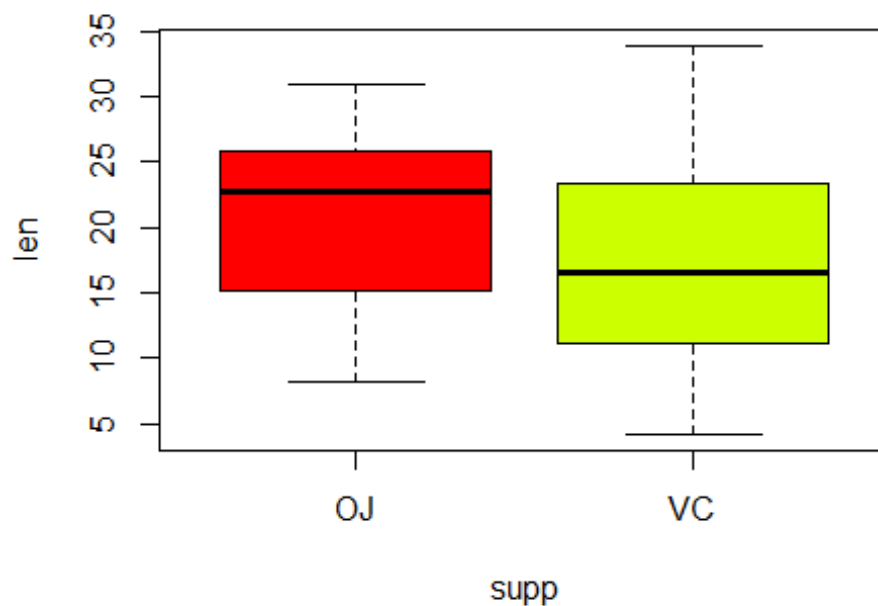
## The following objects are masked from ToothGrowth (pos = 9):
##
##      dose, len, supp

#Question: Does all the groups have the same population mean and use a 5% significance level.?

##Hypothesis Testing
# null hypothesis H0:  $\mu_{VC0.5} = \mu_{OJ1} = \mu_{VC1} = \mu_{OJ2} = \mu_{VC2} = 0$ 
# alternative hypothesis HA: Not all  $\mu_j$  equal 0.

#ONE WAY Anova

boxplot(len ~ supp, col= rainbow(5))
```



```
modell1 <- aov(len ~ supp )
summary(modell1)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## supp          1    205   205.35   3.668 0.0604 .
## Residuals    58   3247    55.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Observation

#The p value for the one way ANOVA of len by supp is 0.0604 which is greater than 0.05, we fail to reject the null hypothesis that $\mu_1 = \mu_2$ at $\alpha = 0.05$. By itself, supplement type (OJ vs. VC) appears to have no significant impact on tooth length.

##For F-value

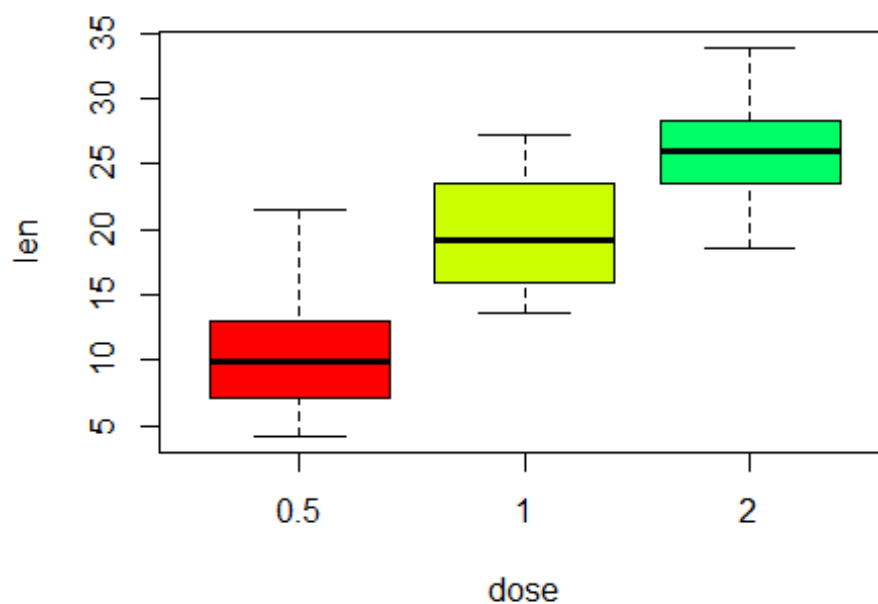
#Finding permuted f-statistics to compare

```
Tobs <- anova(lm(len ~ supp ))[1,4]; Tobs
```

```
## [1] 3.668253
```

#found that the permuted F-statistics, F^ , were less than the observed F-statistic of 3.6682 So fail to reject the null hypothesis*

```
boxplot(len ~ dose, col= rainbow(5))
```



```
model2 <- aov(len ~ dose)
summary(model2)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## dose          1   2224   2224.3    105.1 1.23e-14 ***
## Residuals    58   1228    21.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Observation

*#The p value for the one way ANOVA of len by dose is 1.23e-14 which is less than 0.05, reject the null hypothesis that $\mu_1 = \mu_2$ at $\alpha = 0.05$. means that differences in dosage are associated with different levels of tooth growth
#"larger doses are associated with increased tooth growth.*

##For F-value

#Finding permuted f-statistics to compare

```
Tobs <- anova(lm(len ~ dose, data=ToothGrowth))[1,4]; Tobs
```

```
## [1] 105.0648
```

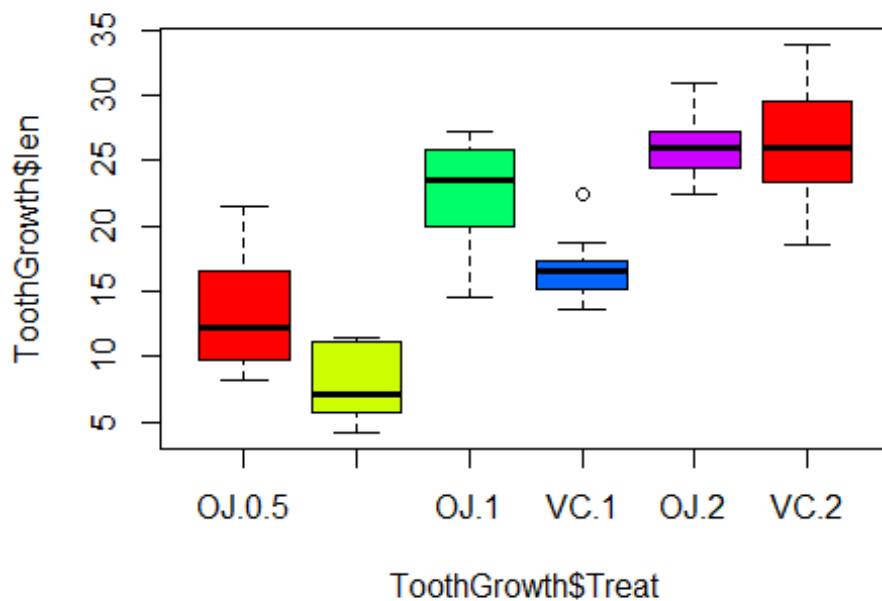
#found that none of the permuted F-statistics, F^ , were larger than the observed F-statistic of 105.06 so reject the null hypothesis*

##Two- Way Anova

#Creates a new variable Treat with 6 levels

```
ToothGrowth$Treat=with(ToothGrowth,interaction(supp,dose))
```

```
boxplot(ToothGrowth$len ~ ToothGrowth$Treat, col= rainbow(5))
```



```
model3 <- aov(ToothGrowth$len ~ ToothGrowth$Treat)
summary(model3)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## ToothGrowth$Treat  5 2740.1   548.0   41.56 <2e-16 ***
## Residuals        54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Observation

*#here is accociation between tooth lenght and tooth treatment
 #Giving guinea pigs orange juice had a greater impact on tooth growth than gi
 ving them ascorbic acid.
 #Reject H_0 since the p-value is less than 5%.The permutation p-value was repo
 rted as 0, means all the group dose not have the same poopulation mean*

##For F-value

#Finding permuted f-statistics to compare

```
Tobs <- anova(lm(len~Treat,data=ToothGrowth))[1,4]; Tobs
```

```
## [1] 41.55718
```

#found that none of the permuted F-statistics, F^ , were Larger than the obser
 ved F-statistic of 41.56 so Reject the null hypothesis*

10. Linear and Multiple Regression Models

```
attach(ToothGrowth)

## The following objects are masked from ToothGrowth (pos = 3):
##
##      dose, len, supp, Treat

## The following objects are masked from ToothGrowth (pos = 10):
##
##      dose, len, supp

##Linear Regression
lm_1 = lm(len ~ dose, ToothGrowth)
summary(lm_1)

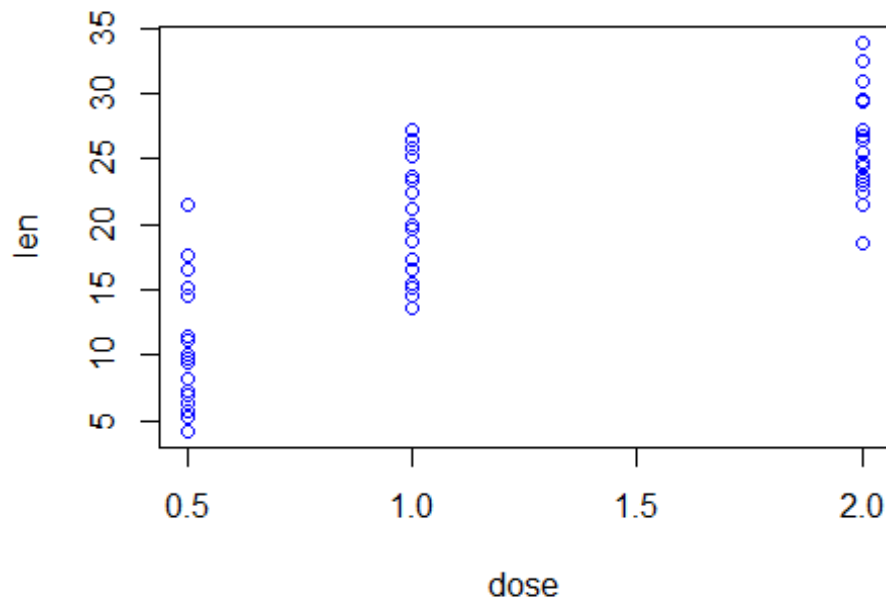
##
## Call:
## lm(formula = len ~ dose, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4496 -2.7406 -0.7452  2.8344 10.1139
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.4225     1.2601    5.89 2.06e-07 ***
## dose          9.7636     0.9525   10.25 1.23e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.601 on 58 degrees of freedom
## Multiple R-squared:  0.6443, Adjusted R-squared:  0.6382
## F-statistic: 105.1 on 1 and 58 DF, p-value: 1.233e-14

##Explanation

# lm name of the model
# p- value for intercept is less than 0.05 which means intercept is significant with this model
# dose has a very significant effect on tooth length as p value is less than 0.05
# positive relation more dose more length
# e= 4.60
# r squared for correlation
# f stats value showing that the value for regression model is very good (both have same p value )...only one x var and one y variable so both p values are same
```


#3Visualization

```
plot(len ~ dose , col = "Blue")
```



#From the graph we can observe a clear positive correlation between the tooth length and the dose

##Multiple Regression

```
fit <- lm(len ~ dose + supp, data=ToothGrowth)
summary(fit)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2725     1.2824   7.231 1.31e-09 ***
## dose          9.7636     0.8768  11.135 6.31e-16 ***
## suppVC       -3.7000     1.0936  -3.383  0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

#Explanation

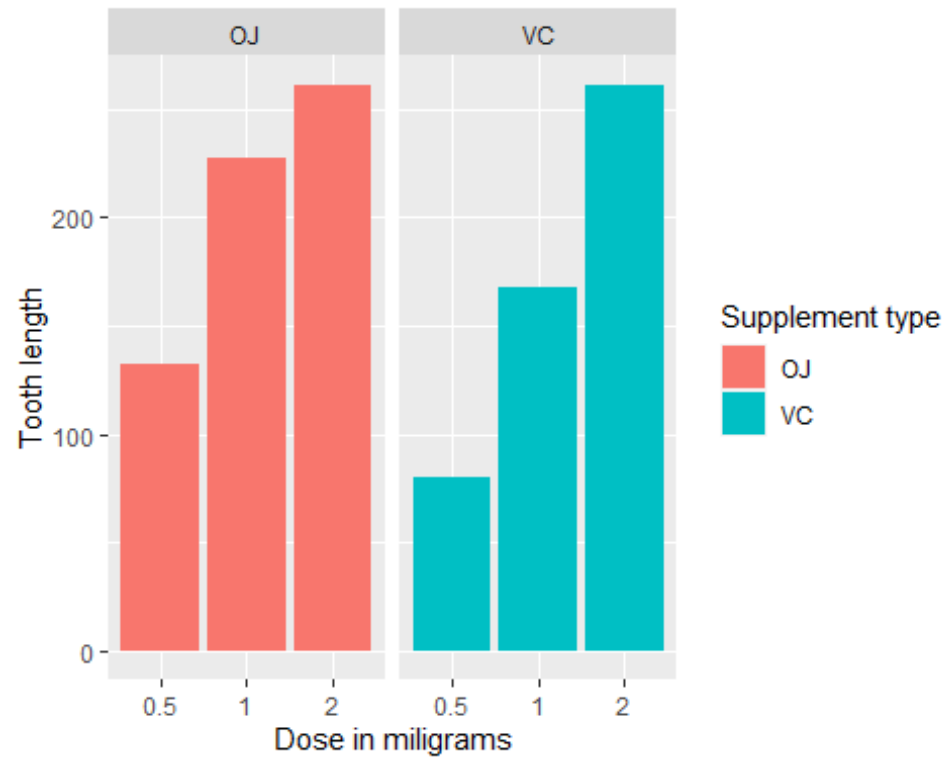
```
#len dependent and dose and supp independent
# lm name of the model
# p- value for intercept is less than 0.05 which means intercept is significant with this model
# supp vc has less significant effect on tooth length as compare to the dose it has more significant effect
# positive relation more dose more length
#negative relation with sup more supp less len
# e= 4.236
# r = 0.83.8 means 84% correlation between depend len all the independent variable
#f stats value showing that the value for regression model is very good (both have same p value )...only one x var and one y variable so both p values are same
```

#The model explains 70% of the variance in the data. The intercept is 9.2725, meaning that with no supplement of Vitamin C, the average tooth length is 9.2725 units. The coefficient of dose is 9.7636. It can be interpreted as increasing the delivered dose 1 mg, all else equal (i.e. no change in the supplement type), would increase the tooth length 9.7636 units. The last coefficient is for the supplement type. Since the supplement type is a categorical variable, dummy variables are used. The computed coefficient is for suppVC and the value is -3.7 meaning that delivering a given dose as ascorbic acid, without changing the dose, would result in 3.7 units of decrease in the tooth length. Since there are only two categories, we can also conclude that on average, delivering the dosage as orange juice would increase the tooth length by 3.7 units.

##Visualization

```
library(ggplot2)
```

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity",) +
  facet_grid(. ~ supp) +
  xlab("Dose in milligrams") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



#From the graph we can observe a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods