Title:
Statistical Inference Class Part 2 - Exploratory Analysis of the ToothGrowth Data – Allen Seol

Basic Summary of the Data:
ToothGrowth Dataset looks at the length growth response in guinea pigs with various different vitamin C doses (0.5,1,2mg) and through orange juice and ascorbic acid. For this project, we will use confidence intervals/hypothesis testing to compare the two tooth groups by supp and dose. Here I look at some summary statistics for the entire dataset as well as different subsets of data.

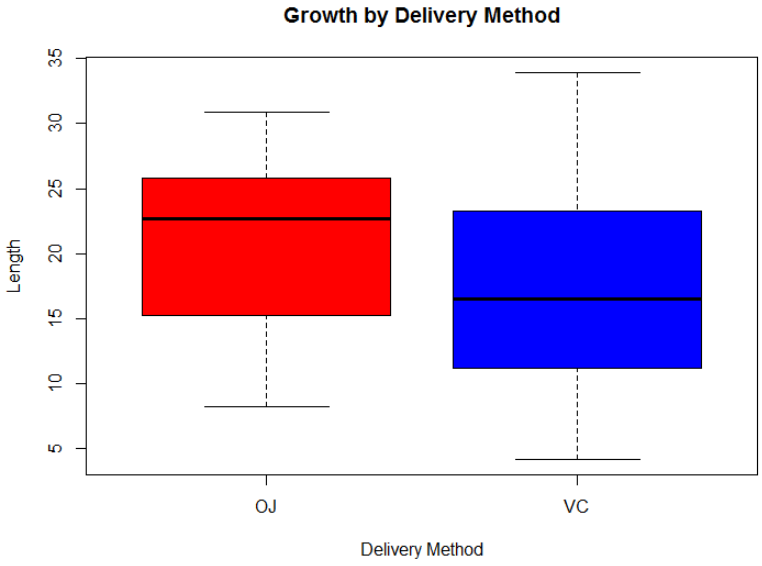
```
##Load DataSet ToothGrowth and Using Dplyr Package
data(ToothGrowth)
library(dplyr)
library(ggplot2)

#subsetting by Dose
dose5 <- filter(ToothGrowth, dose == 0.5 )
dose1 <- filter(ToothGrowth, dose == 1.0 )
dose2 <- filter(ToothGrowth, dose == 2.0 )

#subsetting by Delivery
VC <- filter(ToothGrowth,supp =="VC")
OJ <- filter(ToothGrowth,supp =="OJ")
```

```
> 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

> summary(VC)
      len        supp     dose
 Min.   : 4.20   OJ: 0   0.5:10
 1st Qu.:11.20   VC:30   1  :10
 Median :16.50           2  :10
 Mean   :16.96
 3rd Qu.:23.10
 Max.   :33.90

> summary(OJ)
      len        supp     dose
 Min.   : 8.20   OJ:30   0.5:10
 1st Qu.:15.53   VC: 0   1  :10
 Median :22.70           2  :10
 Mean   :20.66
 3rd Qu.:25.73
 Max.   :30.90
```

```
> summary(dose5)
      len        supp     dose
 Min.   : 4.200   OJ:10   0.5:20
 1st Qu.: 7.225   VC:10   1  : 0
 Median : 9.850           2  : 0
 Mean   :10.605
 3rd Qu.:12.250
 Max.   :21.500
> summary(dose1)
      len        supp     dose
 Min.   :13.60   OJ:10   0.5: 0
 1st Qu.:16.25   VC:10   1  :20
 Median :19.25           2  : 0
 Mean   :19.73
 3rd Qu.:23.38
 Max.   :27.30
> summary(dose2)
      len        supp     dose
 Min.   :18.50   OJ:10   0.5: 0
 1st Qu.:23.52   VC:10   1  : 0
 Median :25.95           2  :20
 Mean   :26.10
 3rd Qu.:27.82
 Max.   :33.90
```

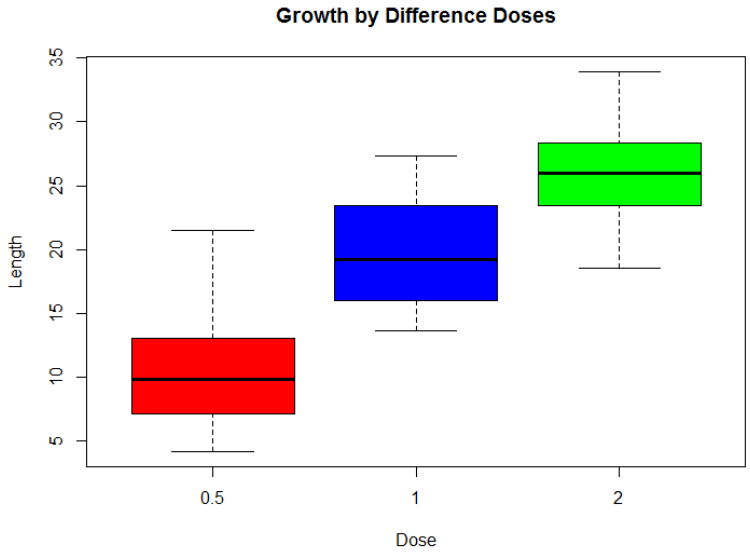
Looking at the VC versus OJ

```
#boxplot OJ vs VC
boxplot(len~supp, data=ToothGrowth, col=(c("red","blue")),
      main = "Growth by Delivery Method", xlab ="Delivery Method", ylab ="Length")
```

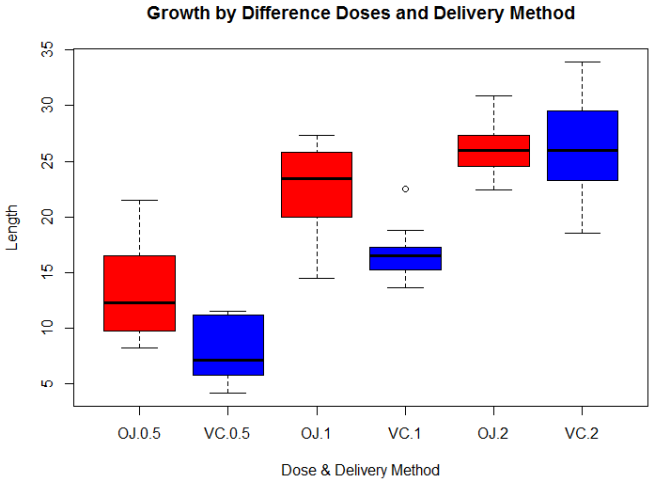


Looking at Dose Groups

```
#boxplot by dose
boxplot(len~dose, data=ToothGrowth, col=(c("red","blue","green")),
      main = "Growth by Difference Doses", xlab ="Dose", ylab ="Length")
```




```
#boxplot by dose and grouping
boxplot(len~supp*dose, data=ToothGrowth,col=(c("red","blue")),
      main = "Growth by Difference Doses and Delivery Method", xlab ="Dose & Delivery Method", ylab ="Length")
```



Hypothesis Testing of Different Supplements and Dosage

```
#T Test by delivery method
t.test(VC$len,OJ$len,paired = FALSE, var.equal = FALSE)
```

```
        Welch Two Sample t-test
data:  VC$len and OJ$len
t = -1.9153, df = 55.309, p-value = 0.06063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.5710156  0.1710156
sample estimates:
mean of x mean of y
 16.96333  20.66333
```

#T Test by Dose Comparison
.5 and 1
t.test(dose5$len,dose1$len,paired = FALSE, var.equal = FALSE)

```
        Welch Two Sample t-test
data:  dose5$len and dose1$len
t = -6.4766, df = 37.986, p-value = 1.268e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.983781  -6.276219
sample estimates:
mean of x mean of y
   10.605    19.735
```

.5 and 2
t.test(dose5$len,dose1$len,paired = FALSE, var.equal = FALSE)

```
        Welch Two Sample t-test

data:  dose5$len and dose1$len
t = -6.4766, df = 37.986, p-value = 1.268e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.983781  -6.276219
sample estimates:
mean of x mean of y
   10.605    19.735
```

1 and 2
t.test(dose1$len,dose2$len,paired = FALSE, var.equal = FALSE)

```
Welch Two Sample t-test
data:  dose1$len and dose2$len
t = -4.9005, df = 37.101, p-value = 1.906e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.996481 -3.733519
sample estimates:
mean of x mean of y
   19.735    26.100
```

Summary and Conclusions

Utilizing both graphical and statistical methods, the delivery method of vitamin C alone shows that the difference between the two means is not significant. However, when you subset the data into subgroups based on the dose of vitamin C given to each group, you can see a very significant difference in means in all possible pair combinations, ie .5 versus 1, 1 versus 2, etc. Look at the growth differences graph based on dosage, there appears to be a dose dependent relationship between length and dose of Vitamin C given.

Based on this data, there is strong evidence to show that the length growth is primarily affected by specifically vitamin C dose rather than the delivery method it self.