Tooth Growth

The ToothGrowth dataset in R contains data on Tooth Growth in 60 guinea pigs where each animal recieved one of three dosages of Vitamin C by one of two methods. Take a look at the data.

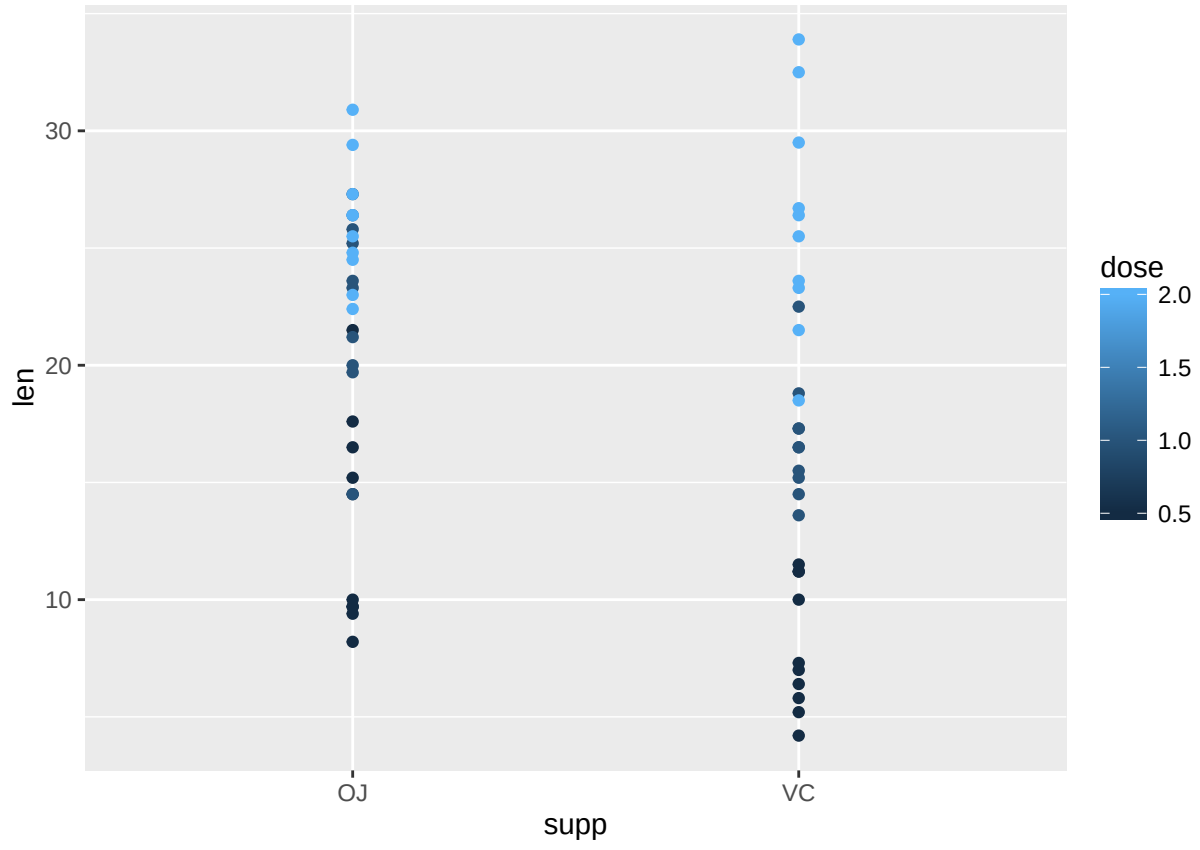
```
library(ggplot2)
str(ToothGrowth)
```

```
## '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 ...
```

```
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
```

```
qplot(data = ToothGrowth, x = supp, y = len,color = dose)
```



So, right away from our exploratory plot we can see that there is a clear relationship between between the dosage amount and the amount of tooth growth. There might also be a relationship between tooth growth and method for providing the dosages but we can't be certain from this graphic.

Method

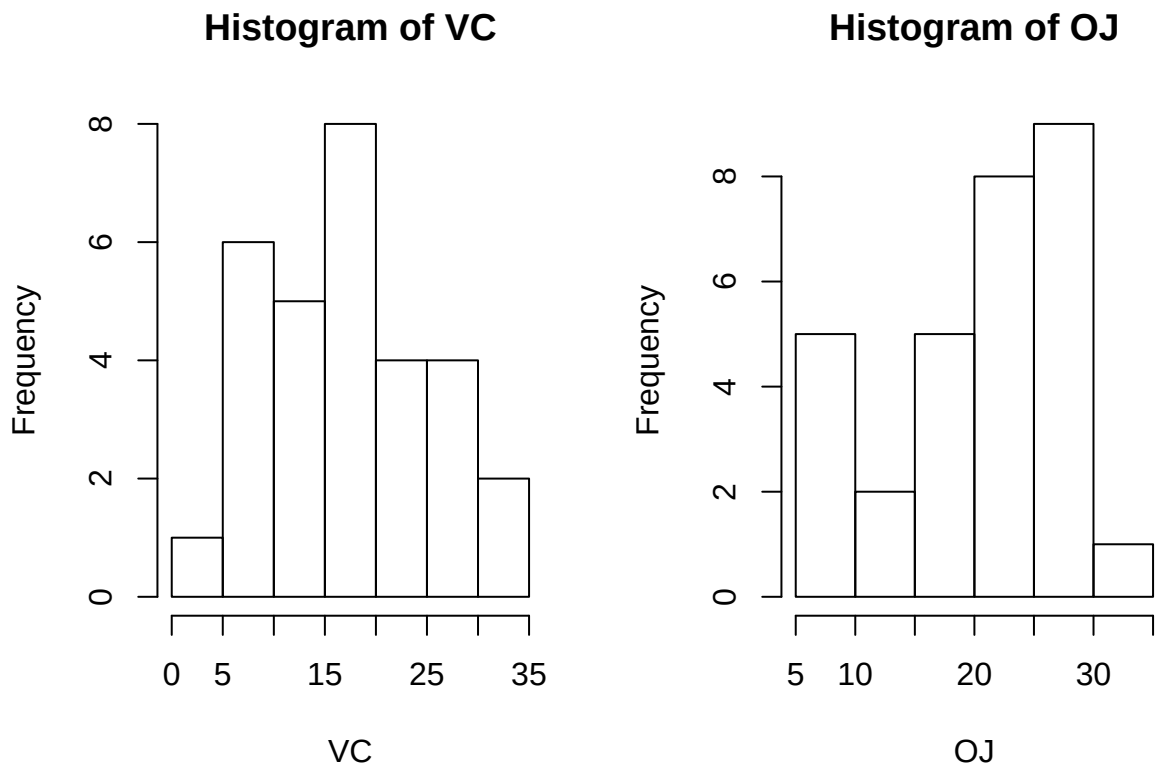
For this, let's dig a little deeper and use confidence intervals in an independent group T test to see if the method chosen has a significant impact on tooth growth.

In order to see conduct the T-test, there are a few robust assumptions that we need to make sure are not unreasonable for our dataset. We first must assume that the guinea pigs chosen for each type of dosage and method were randomly selected in an unbiased manner. We also need our data to be roughly symmetric.

```
table(ToothGrowth$supp)
```

```
##
## OJ VC
## 30 30
```

```
OJ <- ToothGrowth$len[which(ToothGrowth$supp == "OJ")]
VC <- ToothGrowth$len[which(ToothGrowth$supp == "VC")]
par(mfrow = c(1,2))
hist(VC, breaks = 8); hist(OJ, breaks = 8)
```



So, we see that the data is fairly mound shaped and symetric. It could be better, there is some skew with the datapoints under the OJ method but its close enough where we can continue our analysis.

But one thing we can tell from the exploratory plot is that the two groups do not share the same variance. The OJ datapoints look to be much less variable than the VC data points and we can confirm this. Then, we will run an unequal variance T-test to test whether the mean of Tooth Growth for the OJ guinea pigs is statistically different than that of the VC guinea pigs with only allowing 5% occurence of a Type I Error.

```
data.frame( OJ.VAR = var(OJ), VC.VAR = var(VC))
```

```
##     OJ.VAR   VC.VAR
## 1 43.63344 68.32723
```

```
t.test(OJ, VC, var.equal = F)$conf
```

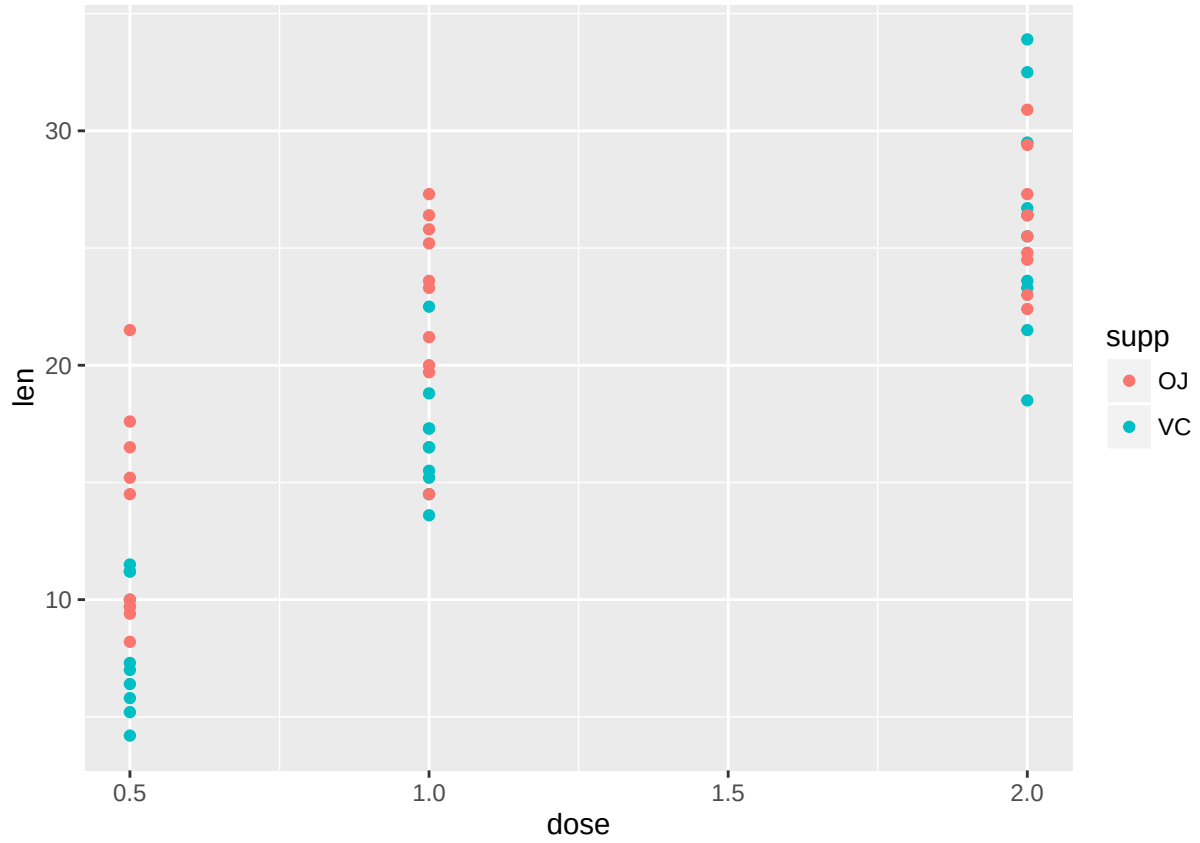
```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

Since our 95% confidence interval of the difference in means contains 0, we will fail to reject the Null Hypthesis that the mean measurements are equal.

Dosages

Now let's look at how Tooth Growth varies based on dosage.

```
qplot(data = ToothGrowth, x = dose, y = len,color = supp)
```



```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

We can see that there is a relationship between dosage amounts and tooth growth. Now, let's walk through and conduct confidence intervals to determine how much growth we can expect with each dosage level.

```
dose.5 <- with(ToothGrowth, len[which(dose == .5)])
dose1 <- with(ToothGrowth, len[which(dose == 1)])
dose2 <- with(ToothGrowth, len[which(dose == 2)])
n <- length(dose.5)

m.5 <- mean(dose.5)
m1 <- mean(dose1)
m2 <- mean(dose2)

sd.5 <- sd(dose.5)
sd1<- sd(dose1)
sd2 <- sd(dose2)

ci.5 <- m.5 + c(-1,1)*qt(.975, n-1)*sd.5/sqrt(n)
ci1 <- m1 + c(-1,1)*qt(.975, n-1)*sd1/sqrt(n)
ci2 <- m2 + c(-1,1)*qt(.975, n-1)*sd2/sqrt(n)
df <- data.frame(Dosage = c(.5, 1, 2))
df <- cbind(df, rbind(ci.5, ci1, ci2))
names(df) <- c("Dosage", "Lower CI", "Upper CI")
df
```

```
##      Dosage  Lower CI Upper CI
## ci.5    0.5  8.499046 12.71095
## ci1     1.0 17.668512 21.80149
## ci2     2.0 24.333643 27.86636
```

So, from the above data frame, we can see about how much tooth growth we would expect with each dosage of Vitamin C. Furthermore, we can conduc inference tests to give a level of confidence for each of our estimates.