project part 2

```r
##So we load the data, and have a look at what the first few lines are like.
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```r
data(ToothGrowth)
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
```

```r
##Then we find the summary which is required.
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  
```

We will then clean the data and look into the variables we need, using the point we are measuring to be the length of the teeth

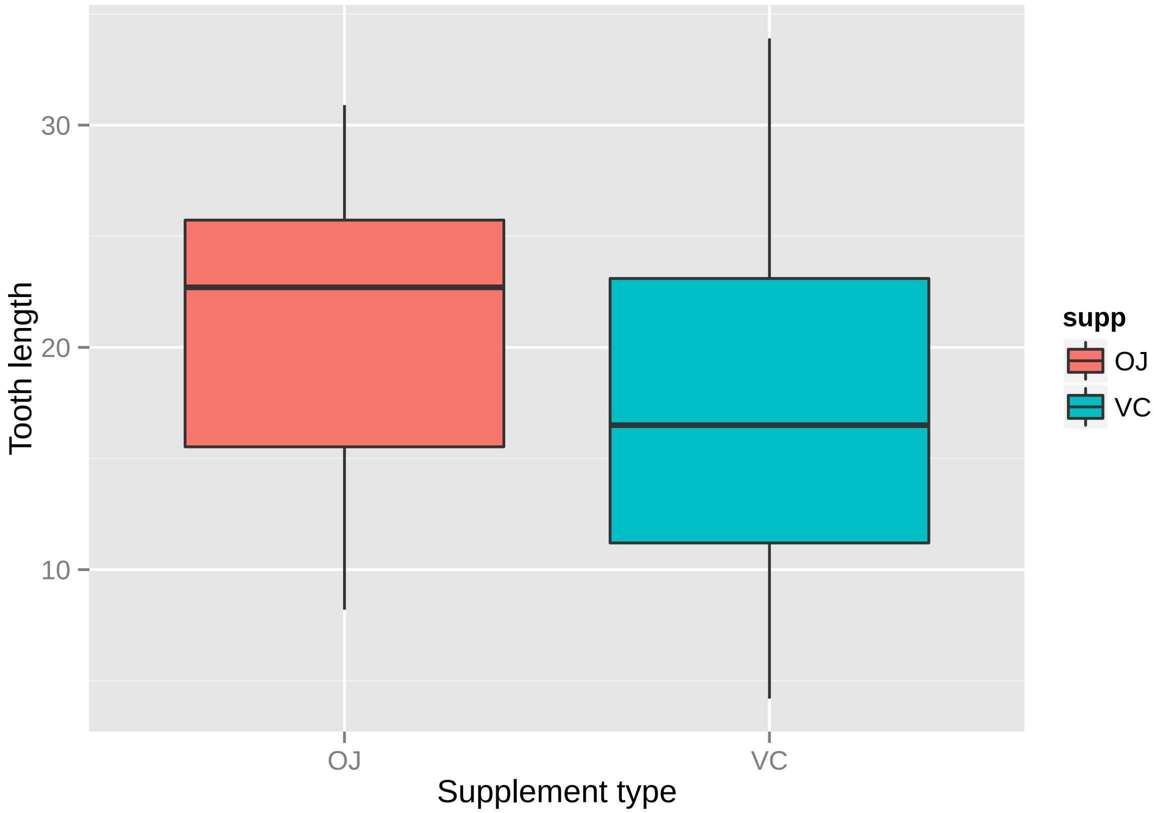
```r
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

MeanSupp = split(ToothGrowth$len, ToothGrowth$supp)
sapply(MeanSupp, mean)
```

```
##       OJ       VC 
## 20.66333 16.96333
```

Then we have a look at the effect of the supplements on the tooth length

```r
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
        xlab("Supplement type") +ylab("Tooth length")
```

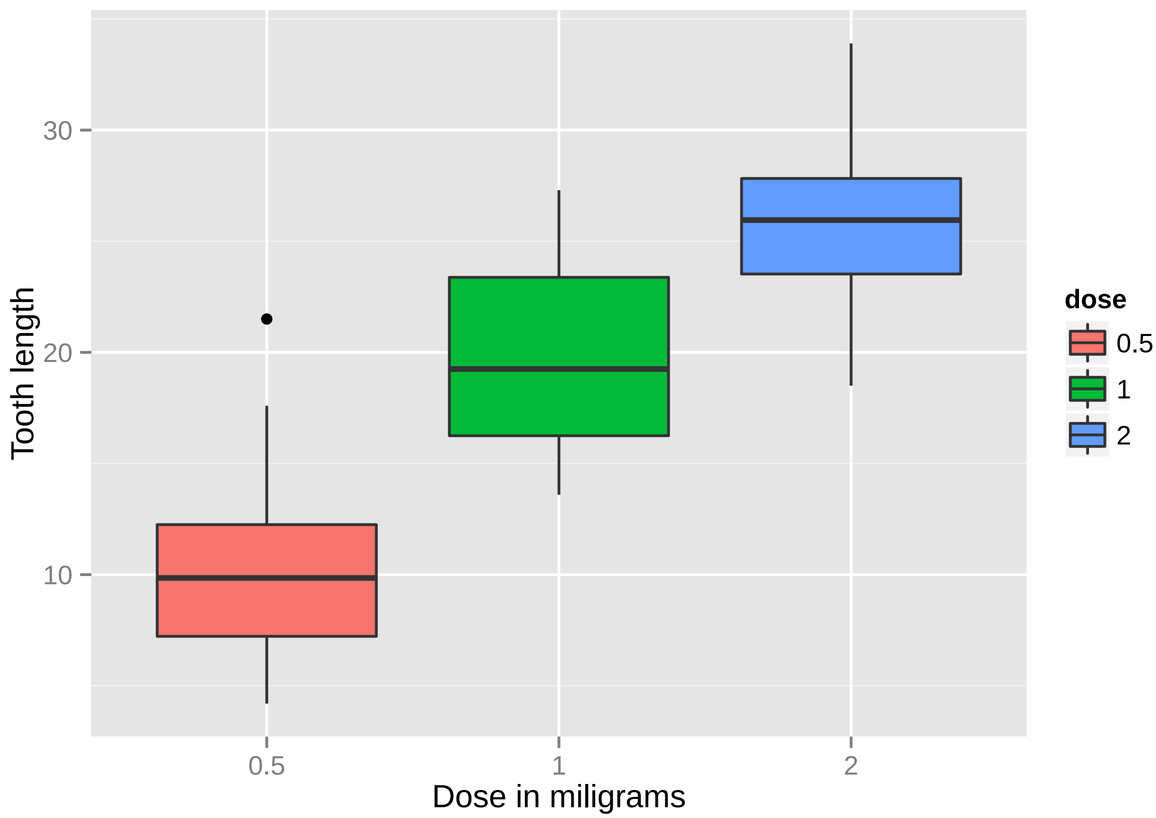


From this, we can conclude that the OJ causes the teeth of guinea pigs to be longer than simply taking in Vitamin C. But what are the effects of different doses of the vitamins?

```
MeanDose = split(ToothGrowth$len, ToothGrowth$dose)
sapply(MeanDose, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
        xlab("Dose in miligrams") +ylab("Tooth length")
```



From the plot, we can see that as the dose gets larger (to 2) the length of the teeth also gets larger - I wonder what other health effects this might have on the poor creatures.

```
len<-ToothGrowth$len
supp<-ToothGrowth$supp
dose<-ToothGrowth$dose

sapply(MeanSupp, var)
```

```
##       OJ       VC
## 43.63344 68.32723
```

```
t.test(len[supp=="OJ"], len[supp=="VC"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[supp == "VC"]
## 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:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The p-value is 6%, which is just slightly larger than the significant value of 5%, which means we can theoretically accept the null hypothesis, but should keep in mind that the confidence level or significant value is only a gauge and such a close fit might require more samples to lean towards a fairer analysis.

```
t.test(len[dose==2], len[dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len[dose == 2] and len[dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The p-value of this test is small enough to be counted as zero, which is evidence to reject the null hypothesis. Which means that the change in Vitamin C dosage is conclusively related to the change in tooth length of guinea pigs.

Using the confidence level of 95%, we will assume that the type of supplements may not have significant impact on the tooth length of guinea pigs, but the increased dosage of Vitamin C has a real impact.