ToothGrowth Data Analysis

From data summary and counting below, we see that the ToothGrowth data has 3 columns. The supp column has 2 values, each with 30 observations.

```
library(datasets)
library(ggplot2)
library(grid)
library(gridExtra)
data(ToothGrowth)
summary(ToothGrowth)
```

```
##      len          supp          dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

The dose column has 3 values, each with 20 observations.

```
## [1] 0.5 1.0 2.0
```

```
## [1] 20 20 20
```

Plot below suggests that as dose increases, tooth lengths also increase for each supplement. It is hard to tell from the plot about the relationship between tooth lengths and supplements.
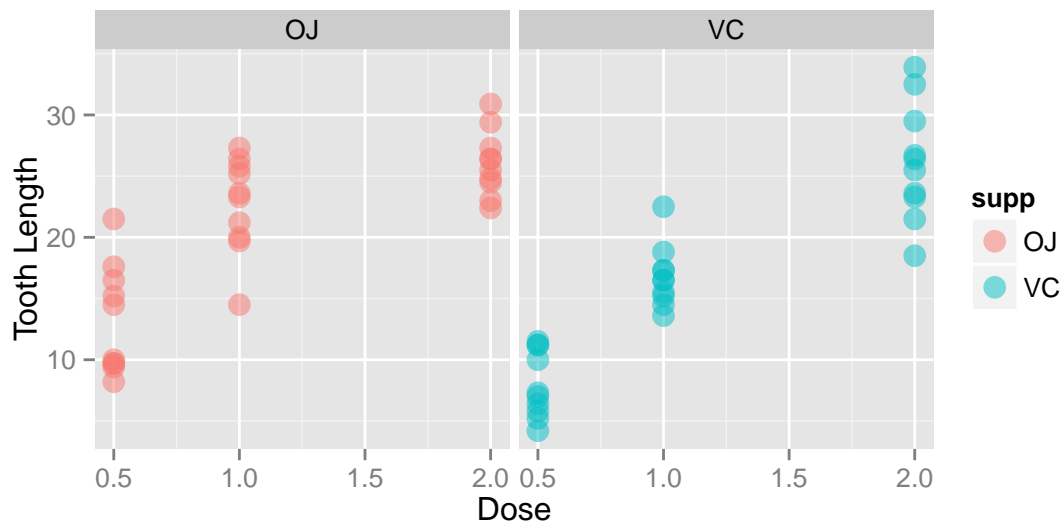
The tooth lengths variance for each dose (0.5, 1.0, 2.0) are as follows:

```
## [1] 20.25 19.50 14.24
```

The tooth lengths variance for each supplement ("OJ", "VC") are as follows:

```
## [1] 43.63 68.33
```

We see that the variances for different doses appear to be different from one another, also the variances for different supplements appear to be different from one other.

- Not knowing the significance of their differences, let us assume that the variances are different.
- Assume that there is no overlap between dose and supplement groups, and that dose level and supplement were randomly assigned to test subjects. We use the unpaired t-test.
- Sample sizes are 20 and 30 for each does and for each supplement, respectively. Both are bigger than 15, so we can assume that the sample means are normally distributed.

Under assumptions above, we perform unequal variance independent T test.

- Comparing the difference in means for the two supplements, we obtain a p-value greater than the significance level 0.05. Therefore, there is not enough difference in the samples to conclude a difference in means. Moreover, the 95% confidence interval below contains 0. Thus we can not reject the null hypothesis that the difference in means is equal to 0.

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)$p.value
```

```
## [1] 0.06063
```

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)$conf.int
```

```
## [1] -0.171  7.571
## attr(,"conf.level")
## [1] 0.95
```

- Comparing the difference in means for the 3 doses by comparing two at a time. All three p-values from tests below are very small. We can reject the null hypothesis that the there is no difference in means. This suggests that dose amount is significant to tooth length.

```
##      [,1]                  [,2]                 [,3]
## [1,] "0.5 to 1.0 p-value"  "1.0 to 2.0 p-value"  "0.5 to 2.0 p-value"
## [2,] "1.26830072017385e-07" "1.9064295136718e-05" "4.39752495936323e-14"
```

- Below are confidence intervals for comparing dose 0.5 to 1.0, dose 1.0 to 2.0, and dose 0.5 to 2.0, respectively. None of the intervals contain 0. We are 95% confident that there are differences in means for all three dose levels.

```
## [1] -11.984  -6.276
## attr(,"conf.level")
## [1] 0.95
```

```
## [1] -8.996 -3.734
## attr(,"conf.level")
## [1] 0.95
```

```
## [1] -18.16 -12.83
## attr(,"conf.level")
## [1] 0.95
```