Statistical Inference Project - Tooth Growth by Ola

We're going to analyze the ToothGrowth dataset in the R datasets package.

Dataset name: 'The Effect of Vitamin C on Tooth Growth in Guinea Pigs'.

Description: 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).

Format: A data frame with 60 observations on 3 variables.

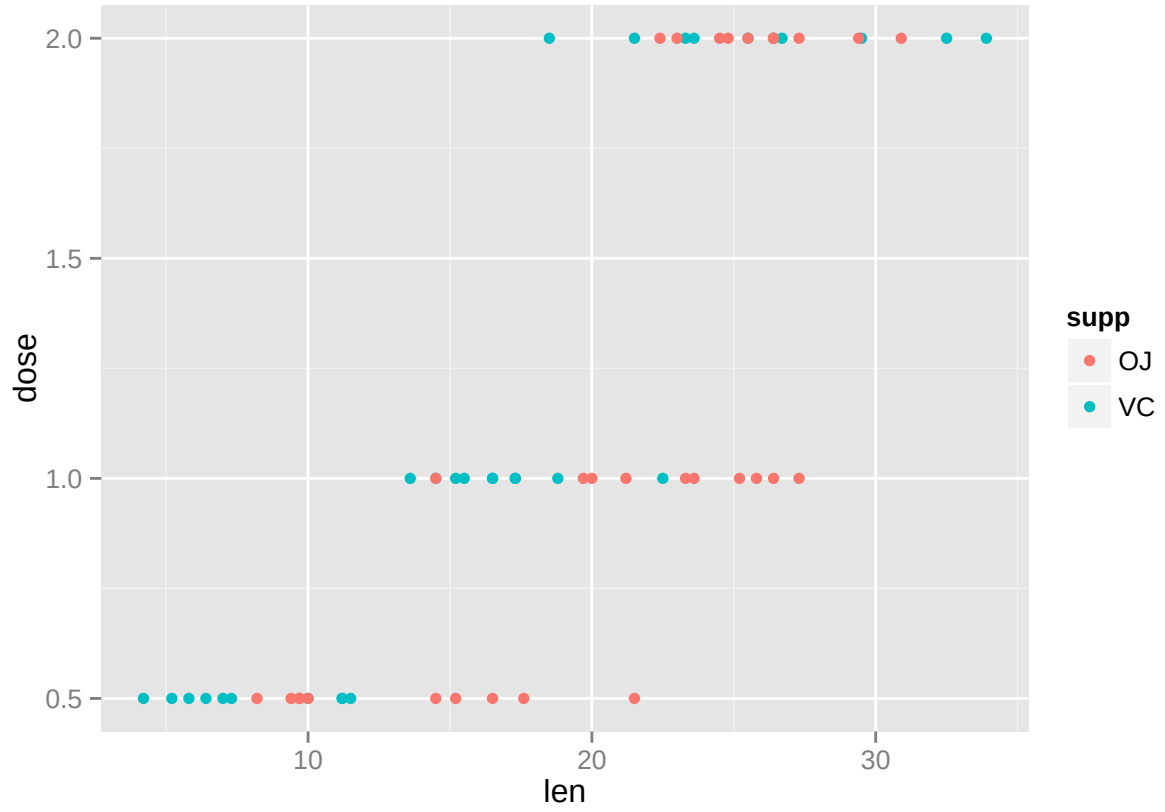
Before I proceed with the project, I add the library ggplot2. I load the data.

```
library(ggplot2)
data(ToothGrowth)
```

1. Load the ToothGrowth data and perform some basic exploratory data analyses

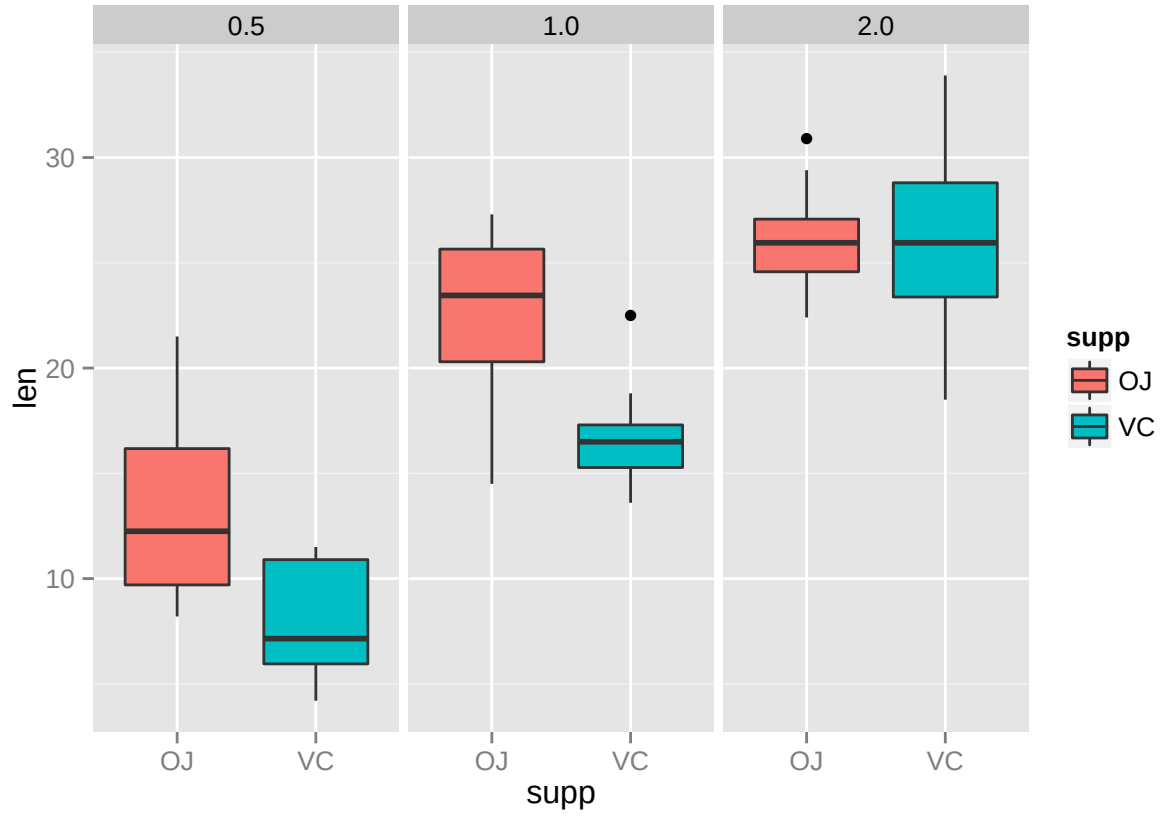
Let's try to show all 3 parameters of our data: length (len), dose (dose) and delivery method (supp). We plot dose vs len and colour our data points due to the variable supp.

```
p1 <- ggplot(aes(x=len, y = dose), data = ToothGrowth)
p1 <- p1 + geom_point(aes(color = supp))
p1
```



This basic plots suggests that there might be a difference in average tooth length for various doses and supp values. Let's see whether this is indeed true:

```
p2 <- ggplot(aes(x = supp, y = len), data = ToothGrowth)
p2 <- p2 + geom_boxplot(aes(fill = supp))
p2 <- p2 + facet_wrap(~ dose)
p2
```



2. Provide a basic summary of the data.

The most basic summary of the data is using the summary function in R.

```
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
```

However, a more appropriate summary would be if we include variablity by dose and supp.

```
by(ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), summary)
```

```
## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : OJ
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.27   16.50   16.77   17.30   22.50
## ------------------------------------------------------------
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.58   25.95   26.06   27.08   30.90
## ------------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

a. Below we test by supp factor and do not consider dosage.

Null hypothesis: there is not a significant difference in tooth length between the two supplement types.

```
testA <- t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
testA$p.value
```

```
## [1] 0.06063451
```

We read the results of the t test as follows: We cannot reject the null hypothesis.

b. Below we test by dosage factor not considering the supplement type. We perform testing between pairs of dosages.

```
Dose12 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Dose13 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Dose23 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

Null hypothesis: increasing the dose level does not lead to an increase in tooth length.

```r
testB1 <- t.test(len ~ dose, paired = F, var.equal = F, data = Dose12)
testB1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

```r
testB2 <- t.test(len ~ dose, paired = F, var.equal = F, data = Dose23)
testB2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 1 mean in group 2
##          19.735          26.100
```

```r
testB3 <- t.test(len ~ dose, paired = F, var.equal = F, data = Dose13)
testB3
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

For this three pairs of dose levels, the p-value is less than 0.05. This means that we can reject the null hypothesis.

C.

4. State your conclusions and the assumptions needed for your conclusions.

Conclusions:

- Supplement type does not influence tooth length.
- Inreasing the dose level leads to an increase in tooth length.

Assumptions

- The 60 gunue pigs in the study are representative of gunea pigs population. The assignment of guinea pigs to different supplements and doses were random.
- The variances for t test are assumed to be different between two compared groups.