

# Project 1 Part 2

*Kaylee Walsh*

*October 23, 2014*

## Part 2

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

## Load Data

```
library(ggplot2)
```

```
data(ToothGrowth)
```

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
```

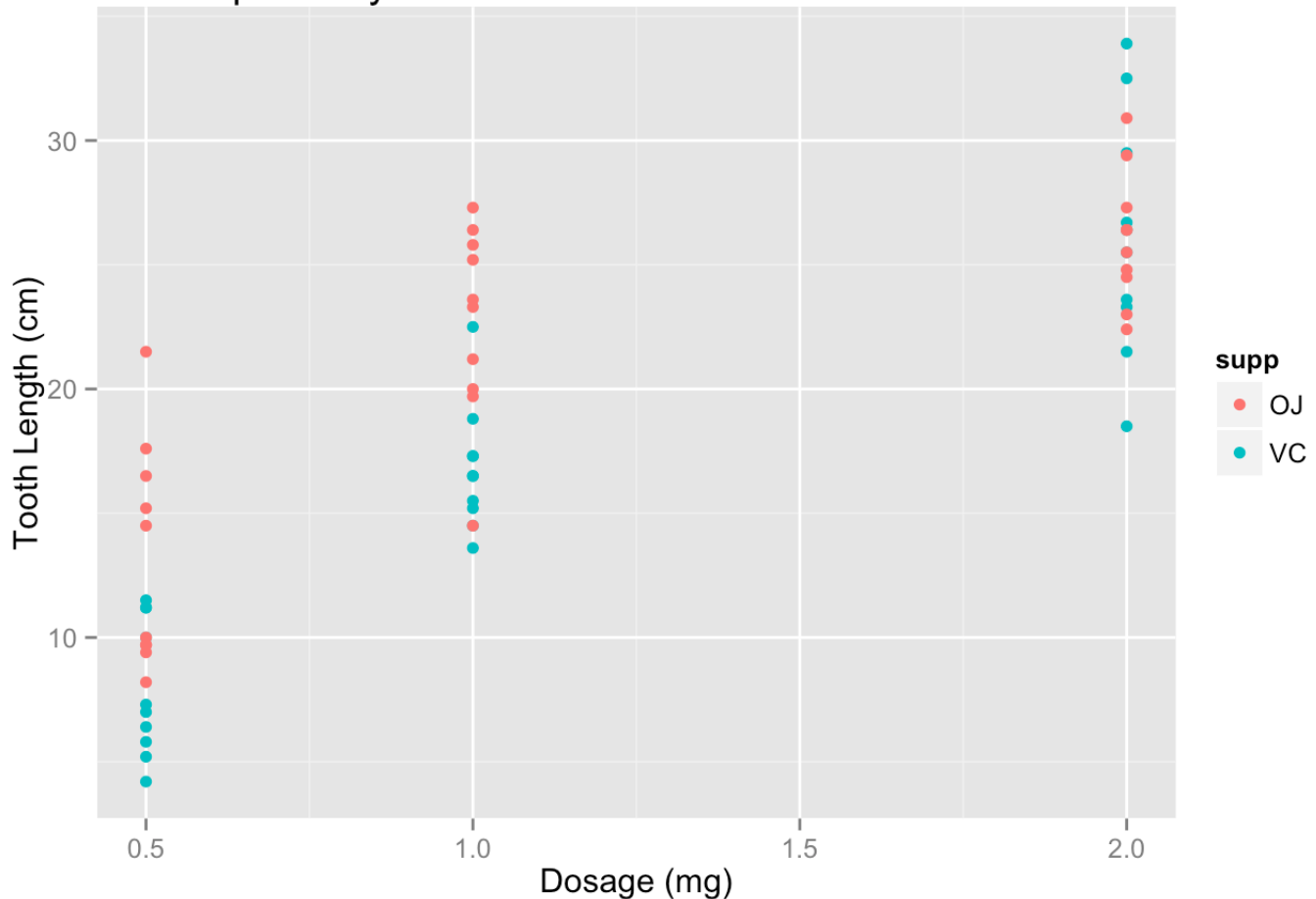
```
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
```

```
## $ 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 ...
```

```
ggplot(x = dose, y = len, data = ToothGrowth, color = supp, main = "Exploratory Visualization  
of the Tooth Growth Data", xlab = "Dosage (mg)", ylab = "Tooth Length (cm)")
```

## Exploratory Visualization of the Tooth Growth Data



## Data Summary

Through the vignette on ToothGrowth we see that the `length` field refers to the response in tooth growth from the dosage of Vitamin C, `dose` in milligrams. The `supp` field corresponds to the supplement type (OJ or VC). These methods were performed on 10 guinea pigs at each dose level and supplement type.

By the plot above we can see that at lowest dose, the orange juice method tends to be more effective than Vitamin C but at highest dose, the two methods are comparably effective.

## Comparing Tooth Growth Methods

Since our sample size is small and we wish to compare the same group under two different methods, we will compare methods by doing a t test. With this test we ask if there is enough difference between the means to show significant difference in efficacy. If the t confidence interval contains 0 then we cannot conclude the two methods are significantly different.

```
g1 <- ToothGrowth[ToothGrowth$supp == 'OJ', 'len']
g2 <- ToothGrowth[ToothGrowth$supp == 'VC', 'len']

difference <- g2 - g1
tmn <- mean(difference)
tsd <- sd(difference)
n <- 30

tmn + c(-1, 1) * qt(0.975, n-1) * tsd / sqrt(n)
```

```
## [1] -5.991 -1.409
```

```
t.test(g2, g1, paired = TRUE)$conf
```

```
## [1] -5.991 -1.409
## attr("conf.level")
## [1] 0.95
```

## Conclusions

From both the t tests we find the confidence interval to not contain 0, thus we can conclude with 95% confidence that the difference in means is not 0 so there is significant different between the two methods of tooth growth.

However to draw these conclusions we assume the distributions are normal and that the two groups are not independent.