

Tooth Growth Analysis

James Sheldon

May 23, 2016

```
sessionInfo()
```

```
## R version 3.3.0 (2016-05-03)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.11.5 (El Capitan)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] ggplot2_2.1.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.5      digest_0.6.9     plyr_1.8.3       grid_3.3.0
## [5] gtable_0.2.0     formatR_1.4      magrittr_1.5     evaluate_0.9
## [9] scales_0.4.0     stringi_1.0-1    rmarkdown_0.9.6  tools_3.3.0
## [13] stringr_1.0.0    munsell_0.4.3    yaml_2.1.13      colorspace_1.2-6
## [17] htmltools_0.3.5  knitr_1.13
```

Overview

This analysis pertains to the ToothGrowth dataset made available in the R platform. We will load the data, summarize what we are seeing, perform some exploratory analysis and use confidence intervals and hypothesis tests to determine any relationships between tooth growth and the other variables.

```
data("ToothGrowth")
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
```

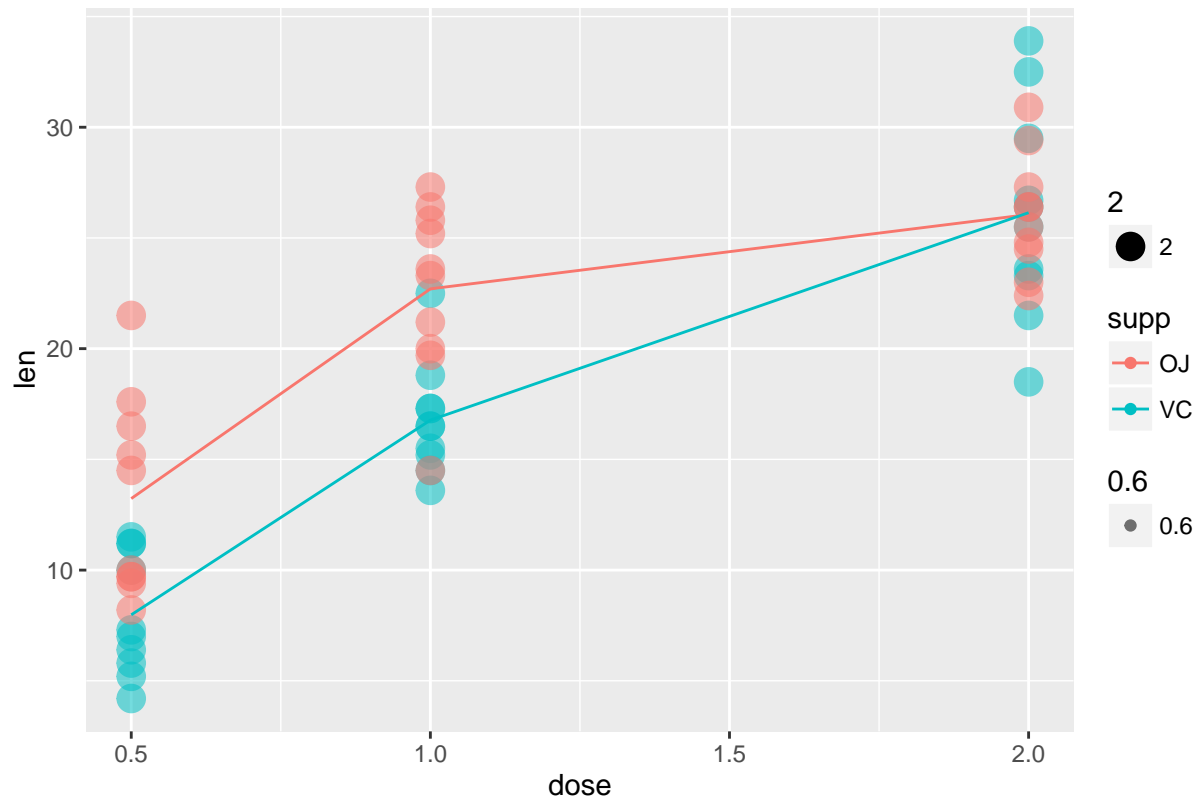
What we're seeing is this data represents the delivery of three different levels of Vitamin C (0.5, 1, and 2 mg) via two different vectors: orange juice or ascorbic acid. There are 60 total observations with three variables:

1. `len` which is a numeric value that represents the tooth length.
2. `supp` is the delivery vector.
3. `dose` is the numeric measure of the dose size in milligrams.

Exploratory Analysis

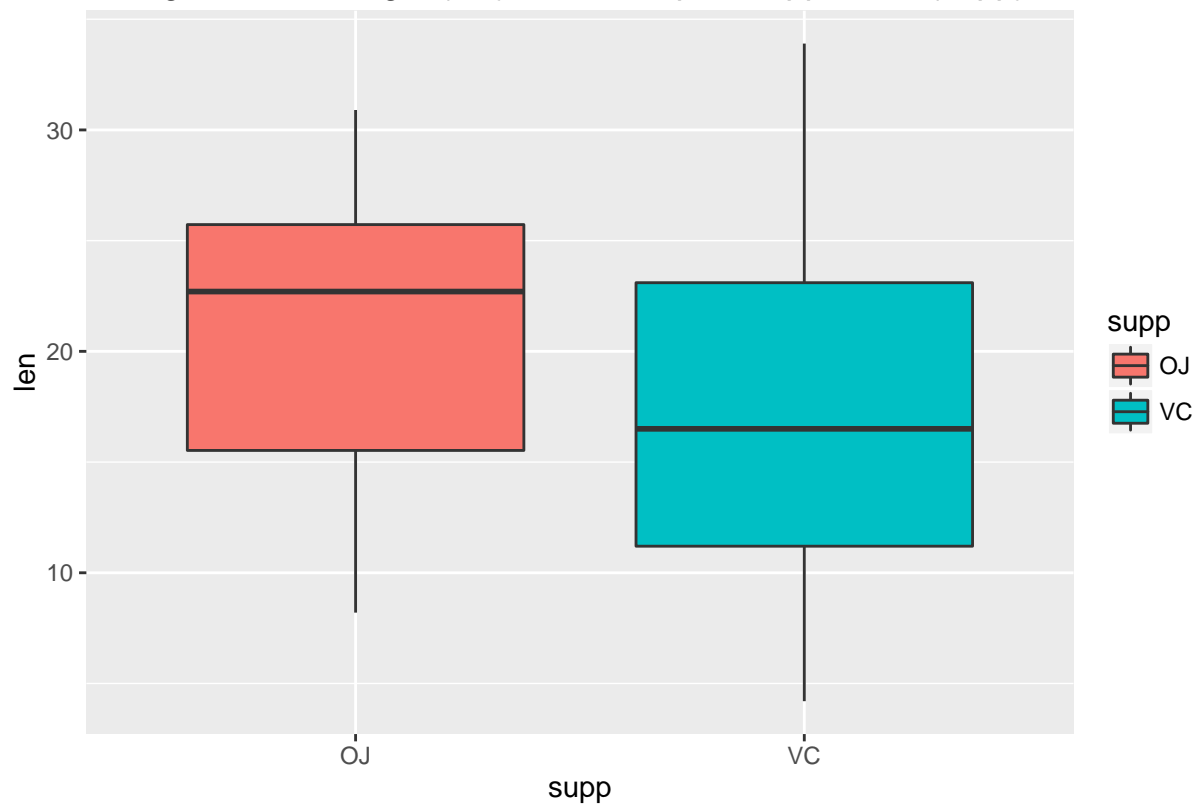
You can also embed plots, for example:

Fig 1: Tooth Length (len) relative to Dosage(dose) & Supplement (supp)



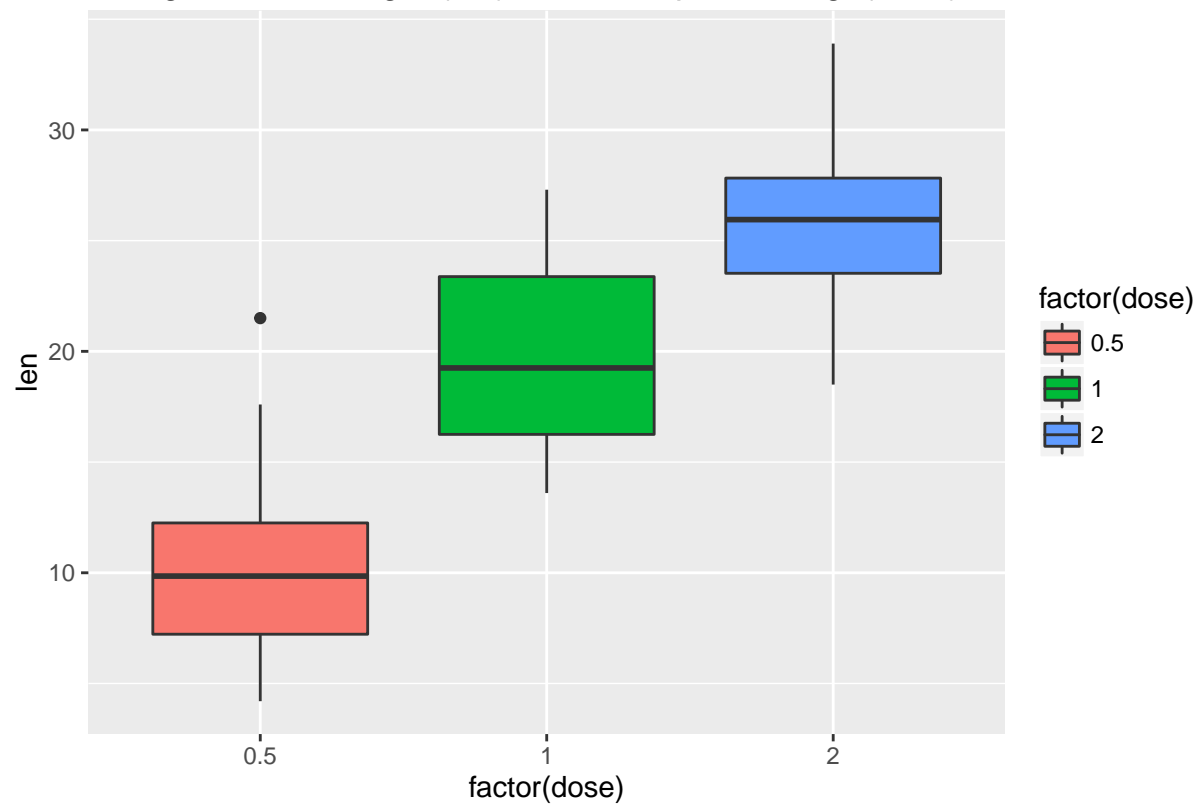
```
# Analyze the Tooth Length (len) relationship to Supplement(supp) independent of dose
g <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
  geom_boxplot(aes(fill = supp))
g <- g + labs(title="Fig 2: Tooth Length (len) relationship to Supplement(supp)")
print(g)
```

Fig 2: Tooth Length (len) relationship to Supplement(supp)



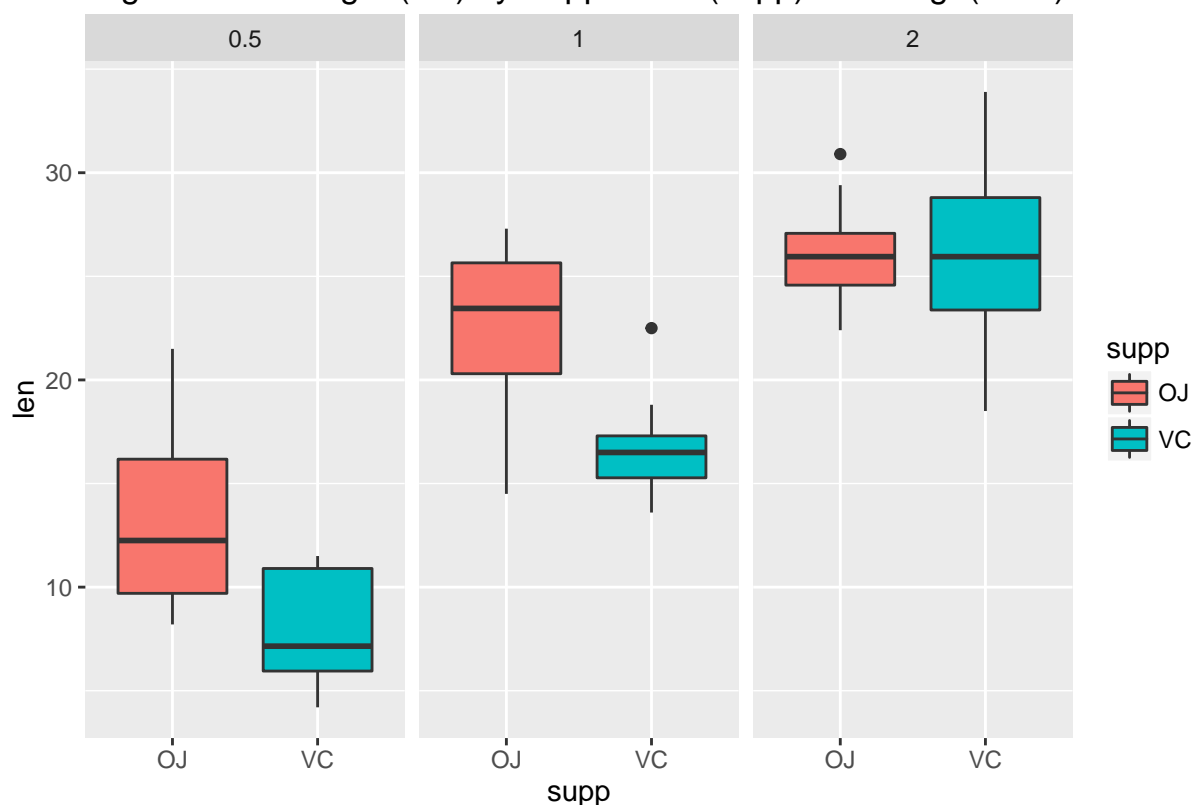
```
# Analyze the Tooth Length (len) relationship to Dose (dose) independent of Supplement
g <- ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
  geom_boxplot(aes(fill = factor(dose)))
g <- g + labs(title="Fig 3: Tooth Length (len) relationship to Dosage(dose)")
print(g)
```

Fig 3: Tooth Lenght (len) relationship to Dosage(dose)



```
# Determine if within a Dosage - supplements have different effects on tooth growth
g <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
  geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose)
g <- g + labs(title="Fig 4: Tooth Lenght (len) by Supplement(supp) & Dosage(dose)")
print(g)
```

Fig 4: Tooth Length (len) by Supplement(supp) & Dosage(dose)



1. Check for group differences due to different supplement type assuming unequal variances between the two groups

```
# Two Sample Test lenght relative to supplement
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## 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 in group OJ mean in group VC
## 20.66333 16.96333
```

```
dose1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = F, var.equal = F, data = dose1)
```

```
##
## 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
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose2)
```

```
##
## 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
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose3)
```

```
##
## 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
```

```
Tooth.dose4 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose5 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose6 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose4)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
## 13.23 7.98
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose5)

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
## 22.70 16.77
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose6)

##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.79807 3.63807
## sample estimates:
## mean in group OJ mean in group VC
## 26.06 26.14
```

The confidence intervals for dose levels 0.5mg and 1.0mg ([1.72, 8.78] within 0.5mg, [2.80, 9.06] within 1.0mg) allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and dose levels. However, the confidence interval for dose level 2.0 [-3.80, 3.64] is not enough to reject the null hypothesis.

5. Conclusions and Assumptions

To make conclusions with the data in this dataset, we must assume that the populations are independent, that the variances between populations are different, a random population was used, the population was comprised of similar guinea pigs, measurement error was accounted for with significant digits, and double blind research methods were used. For the populations to be independent, 60 guinea pigs would have to be used so each combination of dose level and delivery method were not affected by the other methods. To ensure double blind research methods are followed, the researchers taking the measurements must have been unaware of which guinea pigs were given which dose level or delivery method. The guinea pigs must also be unaware that they are being given a specific treatment.

If all the preceding assumptions are true, we may infer that there is a significant difference between tooth length and dose levels across both delivery methods. A higher dose level consistently led to longer teeth. Initially it appeared that the delivery method had no significant impact on tooth length, but when controlling for dose level we discovered that there was a significant difference at 0.5mg and 1.0mg, but not at 2.0mg. Based on this evidence, it appears that orange juice is a better delivery method with a larger impact on tooth length for a given dose of Vitamin C, but above a maximum dose level there is no further improvement.