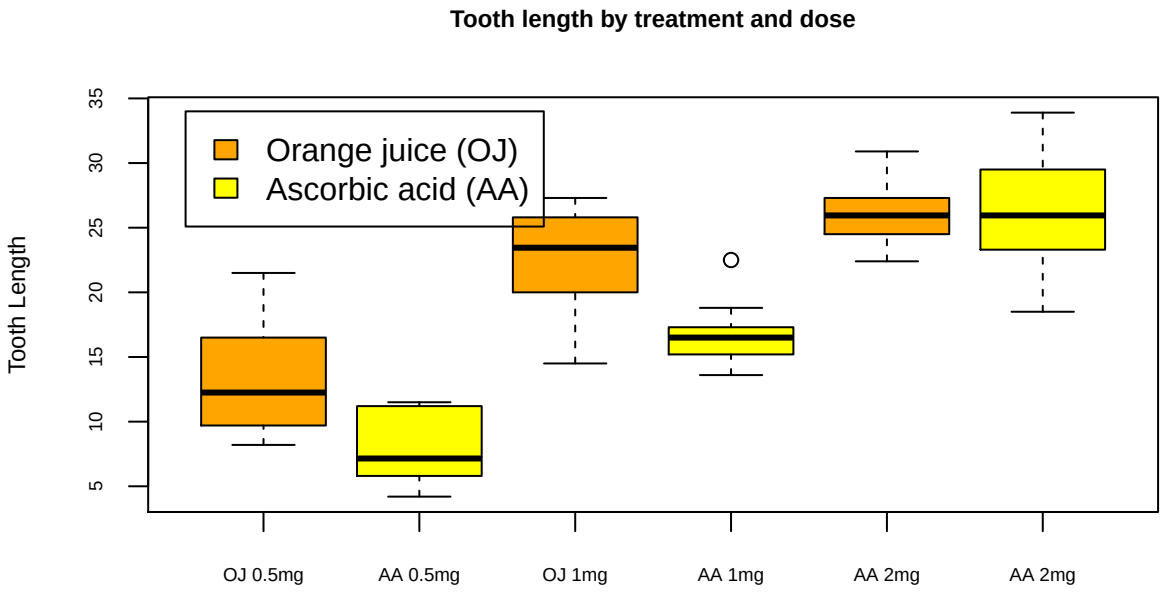
Statistical Inference Project Part 2

1. Looking at the tooth lengths for each treatment condition and dose we can see several things. First, a higher dose of Vitamin C results in longer teeth. Second, for 0.5 mg and 1 mg orange juice results in longer teeth on average than ascorbic acid. However, for 2 mg the average tooth length appears roughly equal between orange juice and ascorbic acid treatments.

```
data(ToothGrowth)
tg <- ToothGrowth
tg$dose <- as.factor(tg$dose)
boxplot(len~interaction(supp,dose),data=tg,ylab="Tooth Length", names=c("OJ 0.5mg","AA 0.5mg","OJ 1mg","AA 1mg","A
legend(0.5, 34, c("Orange juice (OJ)", "Ascorbic acid (AA)"),fill = c("orange", "yellow"))
```



2. Consistent with the boxplots from part 1, the mean tooth length for the 0.5 mg and 1 mg doses of orange juice are higher than for those same doses of ascorbic acid. However, the mean tooth length for 2 mg of orange juice or ascorbic acid are similar. The variance in the data matches what we see in the boxplots in part 1. 0.5 mg and 1 mg of orange juice have a higher standard deviation for tooth length than the corresponding ascorbic acid treatments. However, 2 mg of ascorbic acid has a higher standard deviation for tooth length than the corresponding dose of orange juice.

```
library(dplyr)
```

```
tg.grouped <- group_by(tg,supp,dose)
summarise(tg.grouped, mean(len), sd(len))
```

```
## Source: local data frame [6 x 4]
## Groups: supp
##
##   supp dose mean(len) sd(len)
## 1   OJ  0.5     13.23   4.460
## 2   OJ    1     22.70   3.911
## 3   OJ    2     26.06   2.655
## 4   VC  0.5      7.98   2.747
## 5   VC    1     16.77   2.515
## 6   VC    2     26.14   4.798
```

3. I first want to use a t test to compare whether different doses significantly change mean tooth length.

```
rbind(t.test(tg[tg$dose==0.5, "len"], tg[tg$dose==1, "len"])$conf,
   t.test(tg[tg$dose==0.5, "len"], tg[tg$dose==2, "len"])$conf,
   t.test(tg[tg$dose==1, "len"], tg[tg$dose==2, "len"])$conf)
```

```
##         [,1]    [,2]
## [1,] -11.984  -6.276
## [2,] -18.156 -12.834
## [3,]  -8.996  -3.734
```

```
rbind(t.test(tg[tg$dose==0.5, "len"], tg[tg$dose==1, "len"])$p.value,
   t.test(tg[tg$dose==0.5, "len"], tg[tg$dose==2, "len"])$p.value,
   t.test(tg[tg$dose==1, "len"], tg[tg$dose==2, "len"])$p.value)
```

```
##           [,1]
## [1,] 1.268e-07
## [2,] 4.398e-14
## [3,] 1.906e-05
```

I next want to use a t test to compare whether different supplments significantly change mean tooth length.

```
t.test(tg[tg$supp=="OJ", "len"], tg[tg$supp=="VC", "len"])$conf
```

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

```
t.test(tg[tg$supp=="OJ", "len"], tg[tg$supp=="VC", "len"])$p.value
```

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

I then want to use a t test to compare for a given dose, whether different supplments significantly change mean tooth length.

```
rbind(t.test(tg[tg$supp=="OJ" & tg$dose==0.5, "len"], tg[tg$supp=="VC" & tg$dose==0.5, "len"])$conf,
   t.test(tg[tg$supp=="OJ" & tg$dose==1, "len"], tg[tg$supp=="VC" & tg$dose==1, "len"])$conf,
   t.test(tg[tg$supp=="OJ" & tg$dose==2, "len"], tg[tg$supp=="VC" & tg$dose==2, "len"])$conf)
```

```
##        [,1]  [,2]
## [1,]  1.719 8.781
## [2,]  2.802 9.058
## [3,] -3.798 3.638
```

```
rbind(t.test(tg[tg$supp=="OJ" & tg$dose==0.5, "len"], tg[tg$supp=="VC" & tg$dose==0.5, "len"])$p.value,
   t.test(tg[tg$supp=="OJ" & tg$dose==1, "len"], tg[tg$supp=="VC" & tg$dose==1, "len"])$p.value,
   t.test(tg[tg$supp=="OJ" & tg$dose==2, "len"], tg[tg$supp=="VC" & tg$dose==2, "len"])$p.value)
```

```
##           [,1]
## [1,] 0.006359
## [2,] 0.001038
## [3,] 0.963852
```

4. Conclusions:

- Disregarding the supplement types, higher doses of vitamin c result in longer mean tooth lenghts. I conclude this because the first 3 t tests in part 3 show the 95% confidence interval does not contain 0 and the p value is less than 0.05, so I reject the null hypothesis that there is no difference in mean length.
- Disregarding the doses, there is no difference between ascorbic acid and orange juice because the 95% confidence interval contains 0 and the p value is greater than 0.05, so I do not reject the null hypothesis.
- There is only a significant difference between ascorbic acid and orange juice for 1 mg, because this is the only comparison where the confidence interval does not contain 0 and the p value is less than 0.05

I am making the following assumptions:

- there is not much difference between guina pigs in the different groups except for supplement and dose given.
- that the data is normally distribitated.
- that the data is not paired because it is different guina pigs that receive the different supplments and doses.
- that the variance is not equal between groups based on the boxplots and standard deviations for the different supplments and doses.