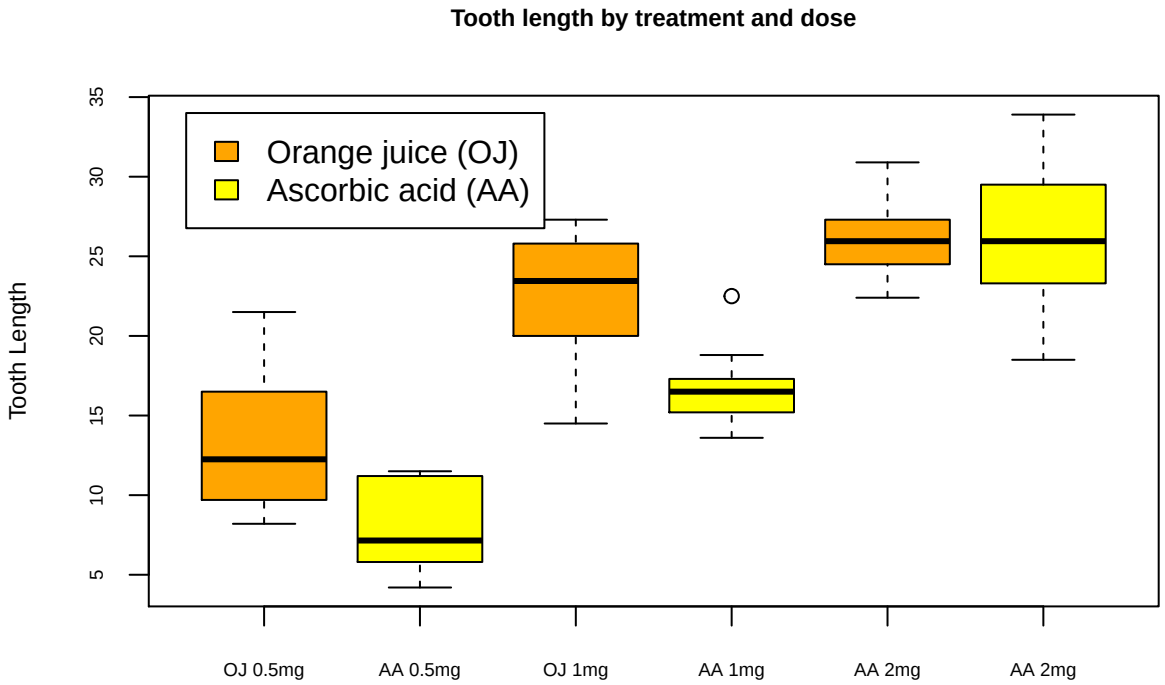
Statistical Inference Project Part 2

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

```
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","AA 2mg","AA 2mg"),
        main="Tooth length by treatment and dose",col=c("orange", "yellow"),
        cex.axis=.6, cex.main=.75, cex.lab=.75)
legend(0.5, 34, c("Orange juice (OJ)", "Ascorbic acid (AA)"),
       fill = c("orange", "yellow"))
```



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.

2. Provide a basic summary of the data.

```
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
```

Consistent with the boxplots from part 1 we see that 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. Additionally, we see that the mean tooth length for 2 mg of orange juice or ascorbic acid are very similar. We also see that 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.

3. t tests between these conditions. Need to do.
4. Assumptions that we make. Don't think it's paired, but the dataset description is confusing. Probably equal variance?