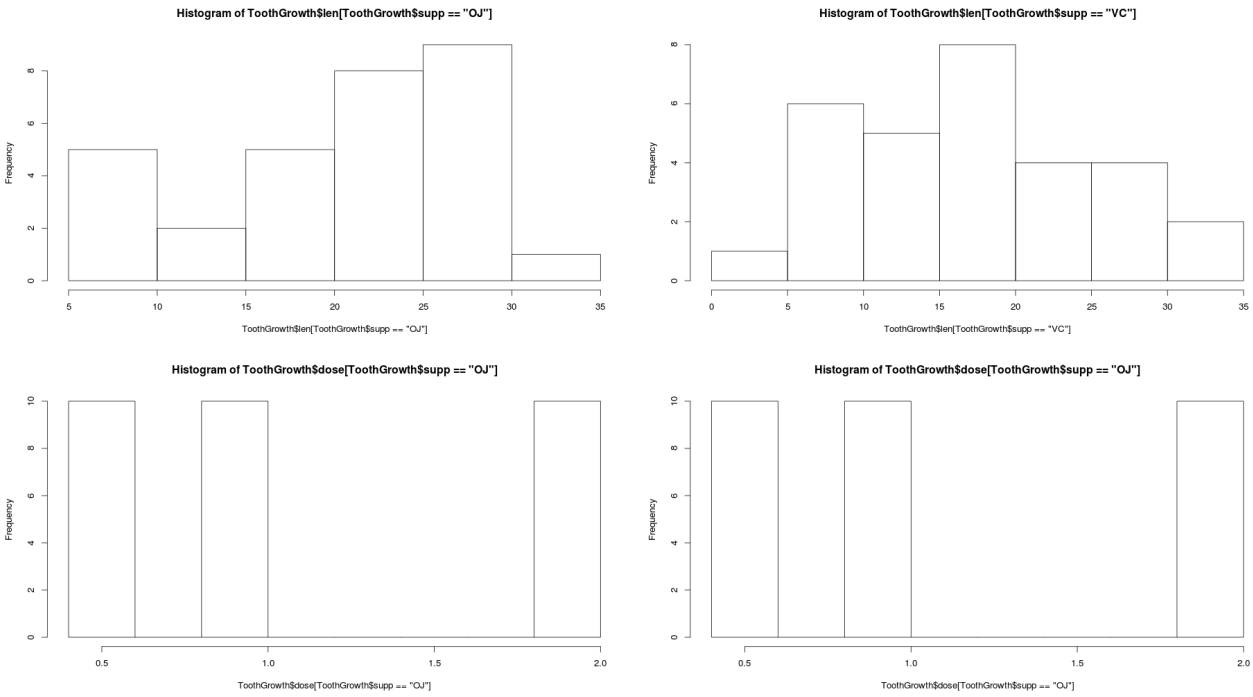
Coursera: Statistical Inference
Project
Q2

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/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)
4. State your conclusions and the assumptions needed for your conclusions.
Some criteria that you will be evaluated on
1. Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
2. Did the student perform some relevant confidence intervals and/or tests?
3. Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
4. Did the student describe the assumptions needed for their conclusions?

1.
```{r}
library(datasets)
data(ToothGrowth)
summary(ToothGrowth)
par(mfrow = c(2,2))
hist(ToothGrowth$len[ToothGrowth$supp=='OJ'])
hist(ToothGrowth$len[ToothGrowth$supp=='VC'])
hist(ToothGrowth$dose[ToothGrowth$supp=='OJ'])
hist(ToothGrowth$dose[ToothGrowth$supp=='OJ'])
```



2.
The summary of the dataset is as follows. And can see from the histogram, the effects of supp on len and dose.

```
      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
```

3.

```{r}
ToothGrowth$supp=as.factor(ToothGrowth$supp)
t.test(len~supp,ToothGrowth)
t.test(dose~supp,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

Welch Two Sample t-test

data:  dose by supp
t = 0, df = 58, p-value = 1
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3278171  0.3278171
sample estimates:
mean in group OJ mean in group VC
       1.166667        1.166667
```

4.
Conlutions
1) we cannot conclude that treatment affects dose
2) we cannot conclude that treatment affects len on a significant level of 95%.
Assumptions
1) observations are random, I.I.D
2) assumed two sample variance differ in this case (default arguments in t.test() function)