Statistical Inference Project, Part 2

First we will load the ToothGrowth built-in dataset that examines the effect of Vitamin C on tooth growth in guinea pigs

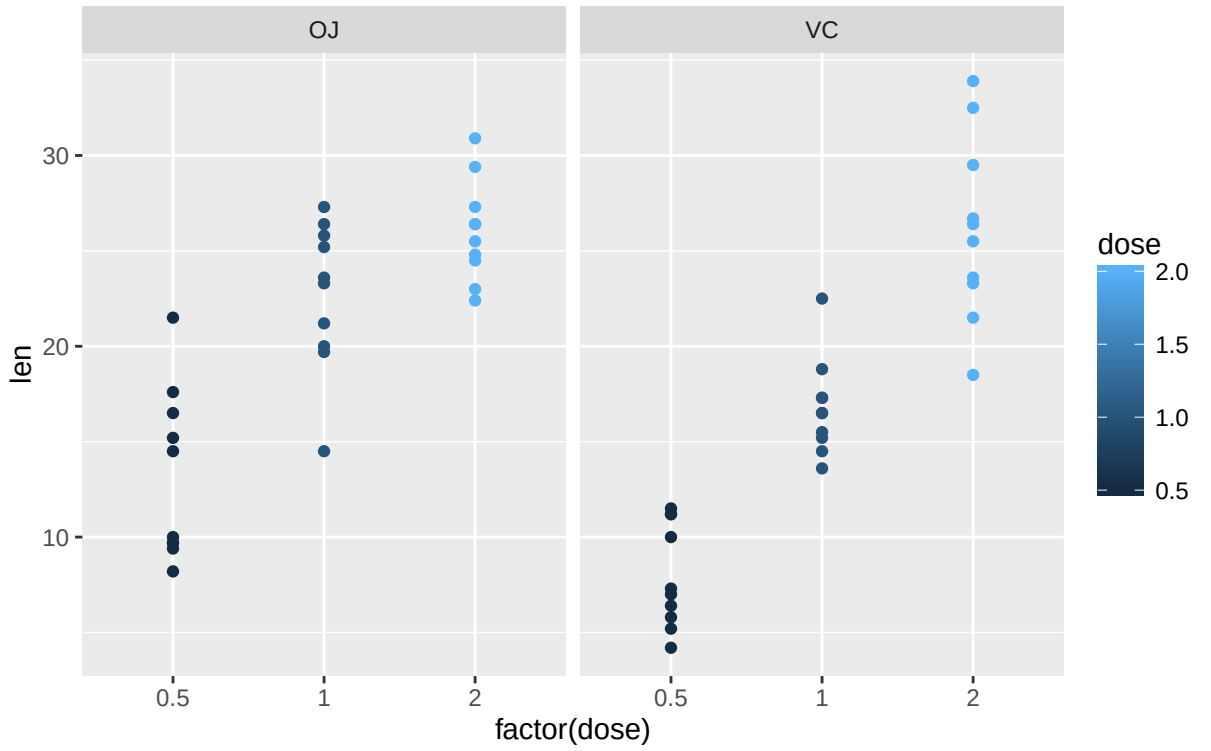
```
data("ToothGrowth")
summary(ToothGrowth)
```

```
##       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
```

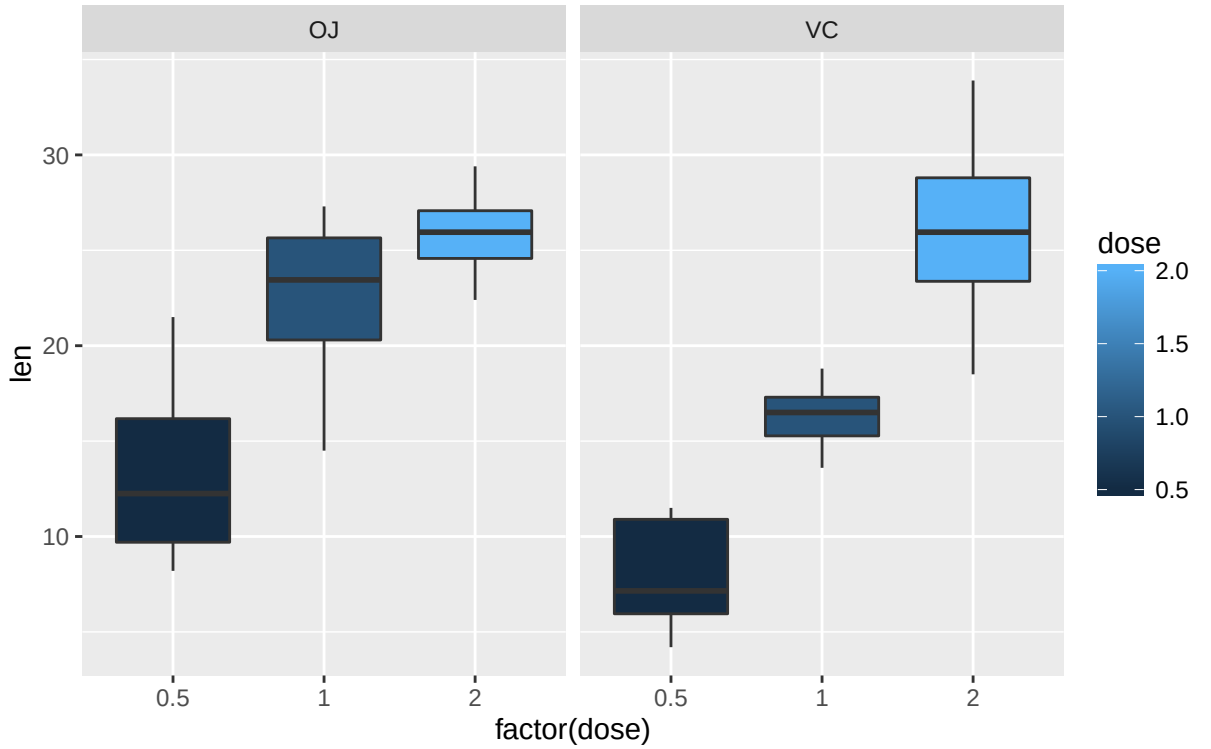
```
library("dplyr")
library("ggplot2")
library("knitr")
grouped = group_by(ToothGrowth,supp,dose)
kable(summarise(grouped,min=min(len),max=max(len),median=median(len),mean=mean(len),sd=sd(len)))
```

supp	dose	min	max	median	mean	sd
OJ	0.5	8.2	21.5	12.25	13.23	4.459708
OJ	1.0	14.5	27.3	23.45	22.70	3.910953
OJ	2.0	22.4	30.9	25.95	26.06	2.655058
VC	0.5	4.2	11.5	7.15	7.98	2.746634
VC	1.0	13.6	22.5	16.50	16.77	2.515309
VC	2.0	18.5	33.9	25.95	26.14	4.797731

```
ggplot(ToothGrowth,aes(x=factor(dose),y=len)) +
  geom_point(aes(color=dose)) +
  facet_wrap(~ supp)
```




```
ggplot(ToothGrowth,aes(x=factor(dose),y=len)) +
  geom_boxplot(outlier.shape = NA,aes(fill=dose)) +
  facet_wrap(~ supp)
```



Confidence intervals for tooth growth by supplement and dose

```
tests = list()
tests[['supp']] = t.test(len~supp, data=ToothGrowth,paired=FALSE)
for (d in c(0.5,1,2)) {
  tests[[toString(d)]] = t.test(len~supp, data=filter(ToothGrowth,dose == d),paired=FALSE)
}
options(digits=2)
kable(sapply(tests,function(t){
  c(t$estimate[1],
    t$estimate[2],
    ci.lower = t$conf.int[1],
    ci.upper = t$conf.int[2],
    p.value  = t$p.value)
}))
```

	supp	0.5	1	2
mean in group OJ	20.66	13.23	22.7	26.06
mean in group VC	16.96	7.98	16.8	26.14
ci.lower	-0.17	1.72	2.8	-3.80
ci.upper	7.57	8.78	9.1	3.64
p.value	0.06	0.01	0.0	0.96

From these results, it appears that regarding supplement type we can not conclude whether the type of supplement at all doses can predict the magnitude of tooth growth since the p-value is close to 0.5 and the confidence interval includes 0 in its range. This can be seen in the graph since the distribution of length as a function of supplement type overlaps for OJ and VC.

Regarding dosage, it can be confirmed that at a dosage of 0.5 or 1.0, supplement type can predict the magnitude of growth. Specifically there could be a link between OJ and greater tooth growth relative to VC. This same conclusion can not be made for a dosage of 2.0 since the p-value is greater than 0.05. The graphs confirm this conclusion since at 0.5 and 1.0 dosages the observed growth range was not overlapping between the supplements.