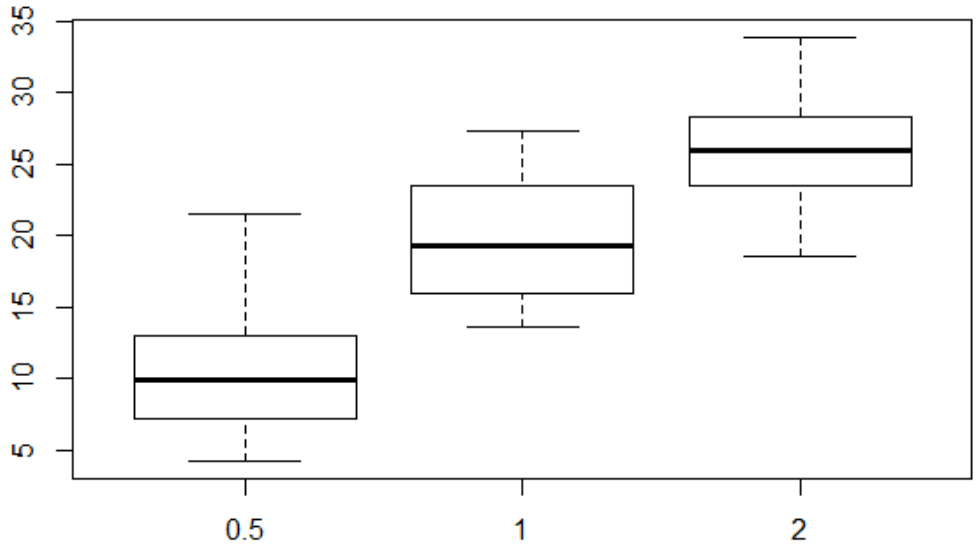
Title：Basic inferential data analysis

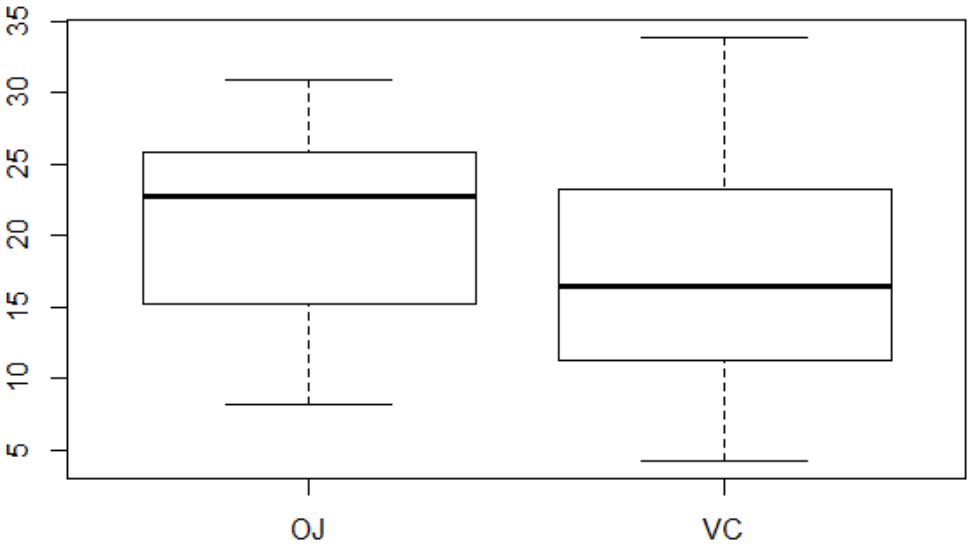
Author：eyangs

Load data and basic exploratory data analyses：

```
attach(ToothGrowth)
boxplot(len~dose)
boxplot(len~supp)
```

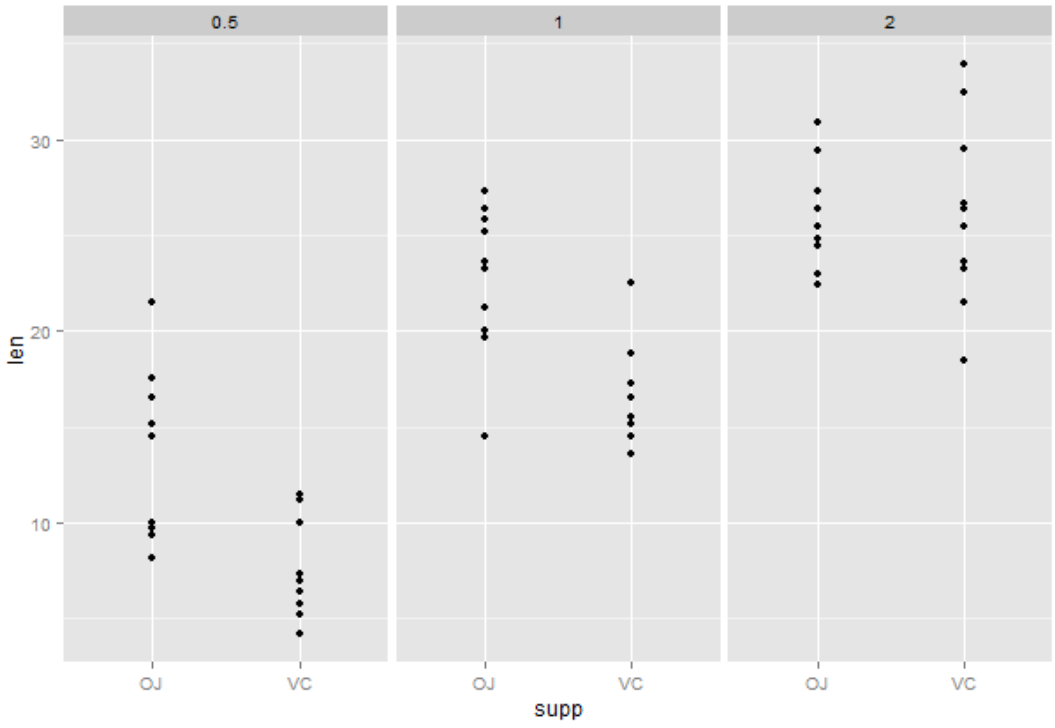


Apparently，mean of len increase with dose. So It is supposed that there is a strong correlation.



#there is no significant correlation between len and sup

```
ggplot(ToothGrowth, aes(supp, len)) + geom_point()+ facet_grid(.~dose)
```



A basic summary of the data：

```
> 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
```

hypothesis tests：(same variance is assumed)

H^0： $\mu^{OJ} = \mu^{VC}$

H^1： $\mu^{OJ} != \mu^{VC}$

#using t.test

t.test(len~supp)

#as the result, H^0 will not be refused. It means that there is no significant difference between en and sup

```
> t.test(len~supp)

 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
```

H^0：mean of len with different dose has same value

H^1：mean of len with different dose has different value

using t.test, p-value is very small, so H^0 should be refused. It means that there is significant correlation between len and dose.

t.test(len,dose)

```
> t.test(len,dose)

 Welch Two Sample t-test

data:  len and dose
t = 17.8096, df = 59.798, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 15.66453 19.62881
sample estimates:
mean of x mean of y
18.813333  1.166667
```