Project 3: Python Project and Performing EDA
Seungchul Yeom / sy9276
5/10/2021
1. Import the data

In [6]:
```
import pandas as pd
ToothGrowth = pd.read_excel(r'C:\Users\Chris\Desktop\UT AUSTIN\Junior spring seme
ToothGrowth.head()
```

Out[6]:

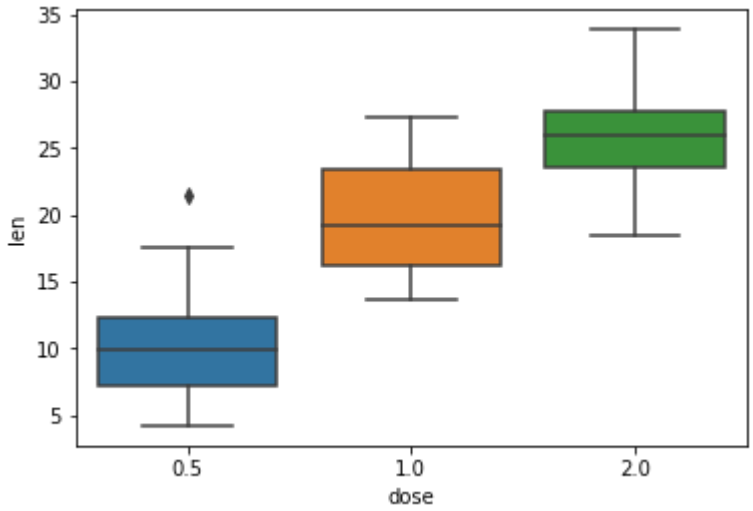
	len	supp	dose
0	4.2	VC	0.5
1	11.5	VC	0.5
2	7.3	VC	0.5
3	5.8	VC	0.5
4	6.4	VC	0.5

2. Describe some features of the data (number of variables / observations)

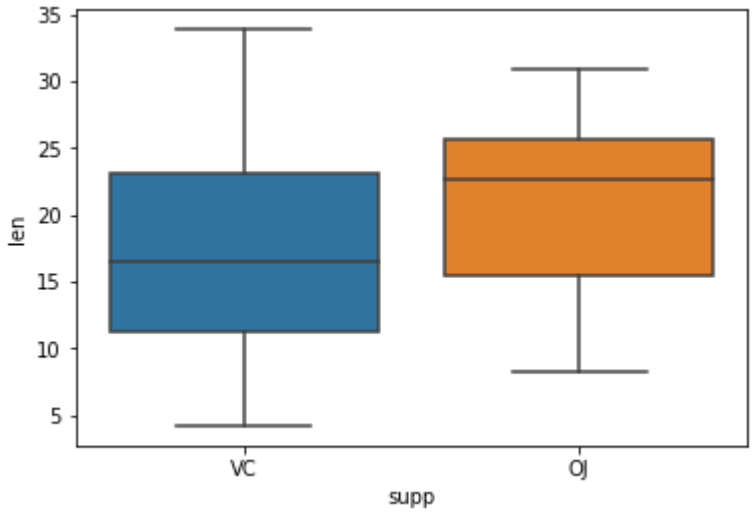
There are 3 variavles which represents Tooth length, supplement type (VC or OJ), and numeric dose in milligrams per day. There are 60 observations, and each observation indicates a single Guinea pigs that take certain dose of vitamin C to observe the effect on tooth growth. The dose levels of vitamin C were 0.5, 1.0, 2.0 mg/day, and the supplement types were VC and OJ which are short stand of Vitamin C and Orange Juice.

3. Perform EDA

In [8]:
```
import seaborn as sb
box = sb.boxplot(x="dose", y = "len", data = ToothGrowth)
```



In [12]:
```
boxsupp = sb.boxplot(x = "supp", y = "len", data = ToothGrowth)
```



In [14]:
```
ToothGrowth.groupby('dose').mean()
```

Out[14]:

dose	len
0.5	10.605
1.0	19.735
2.0	26.100

In [16]:
```
ToothGrowth.groupby('dose').std()
```

Out[16]:

dose	len
0.5	4.499763
1.0	4.415436
2.0	3.774150

Based on the box plot and mean group by dose, the tooth length of Guinea pig seems to increase as amount of vitamin C dose increases. The means of the tooth length for dose 0.5, 1.0, 2.0 are 10.605, 19.735, and 26.100, and the standard deviations of the tooth length for each dose are 4.499, 4.415, and 3.774, respectively.