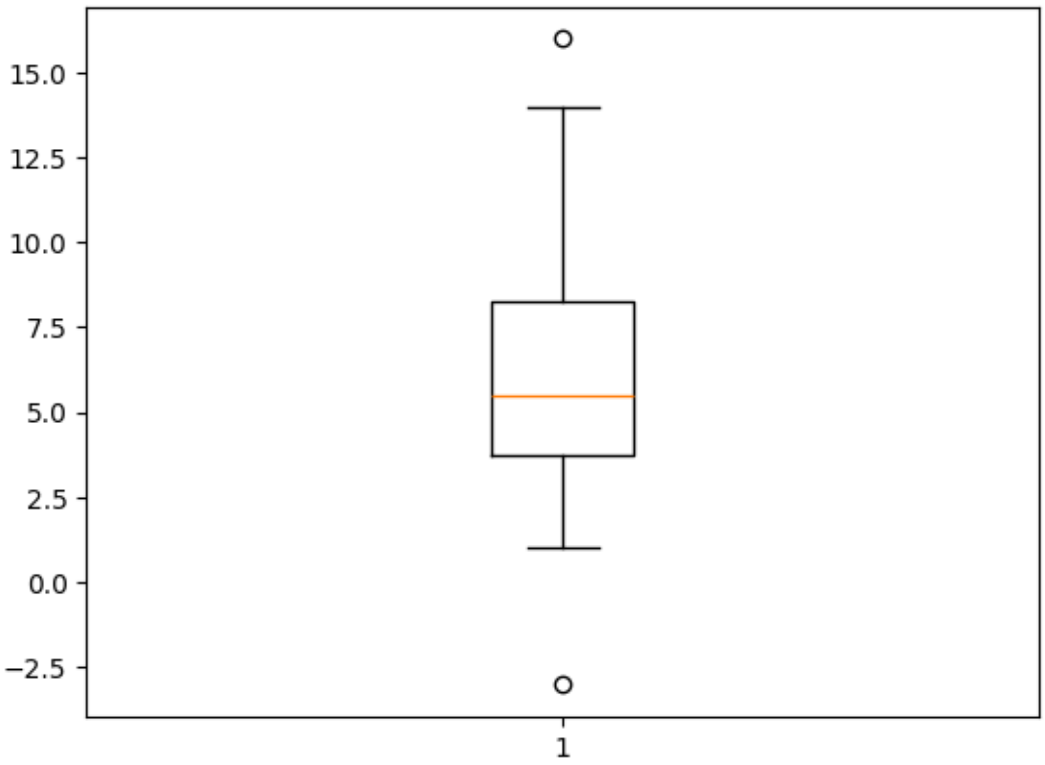
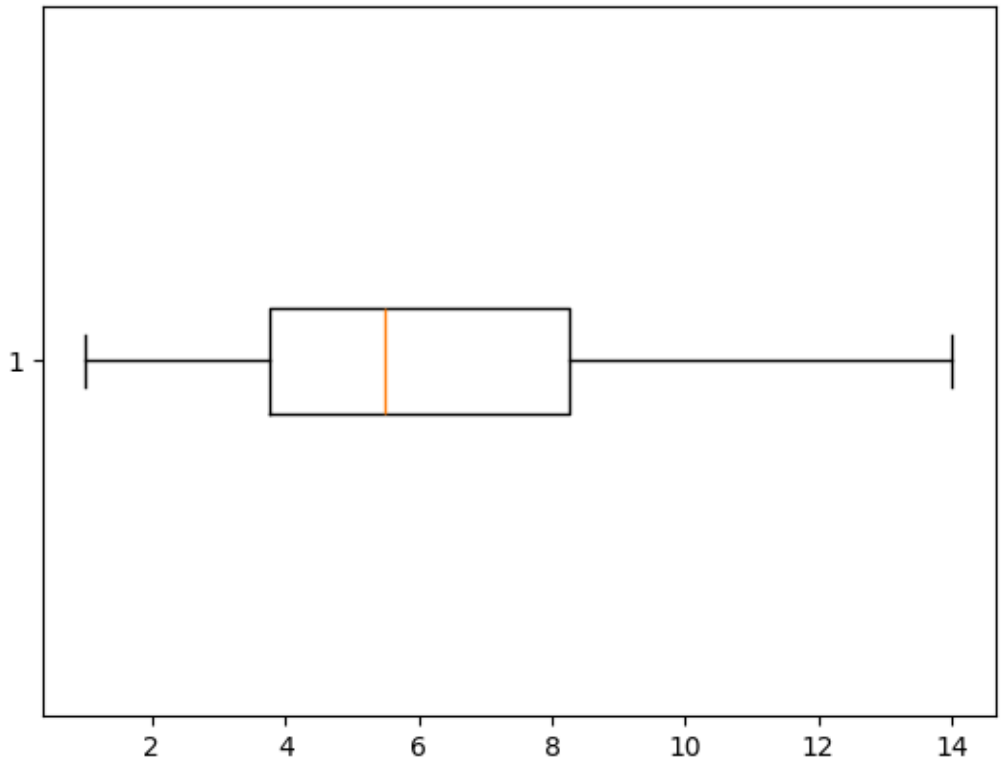
# boxplot

November 1, 2024

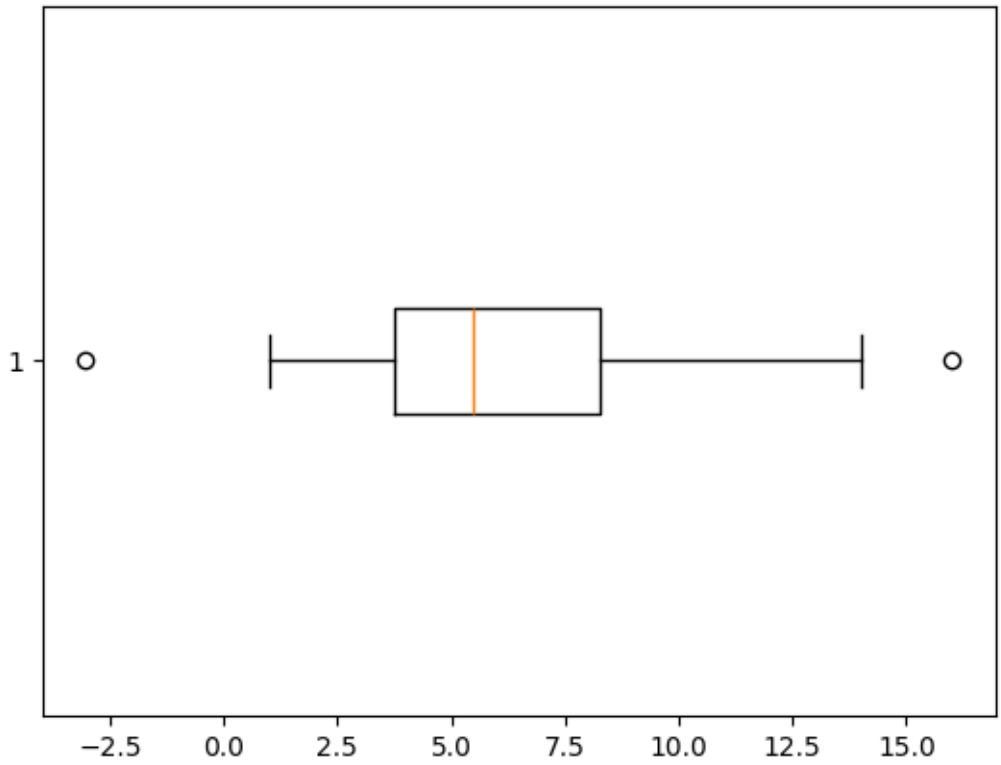
[1]:
```
import matplotlib.pyplot as plt
from matplotlib.pyplot import boxplot, show   #libraries req for boxplot
values = [2, 3, 4, 1, -3.04, 5, 4, 6, 7, 2, 4, 6, 8, 6 , 9, 12, 14, 11, 5, 16]
 ↪ #datapoints need not be ordered
plt.boxplot(values, vert=True,showfliers=True)    #simple way to create a
 ↪boxplot
plt.show()
#aka whisker plot
```



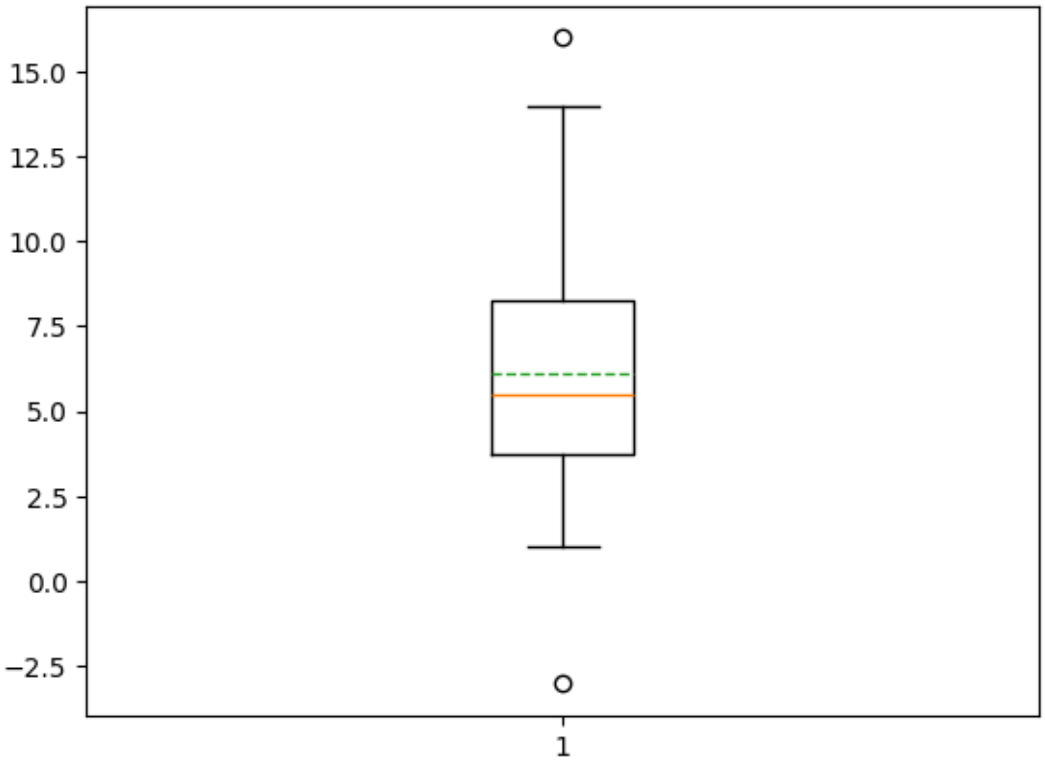

[2]:
```
plt.boxplot(values, showfliers=False, vert=False)   #to remove all the outliers
plt.show()
```



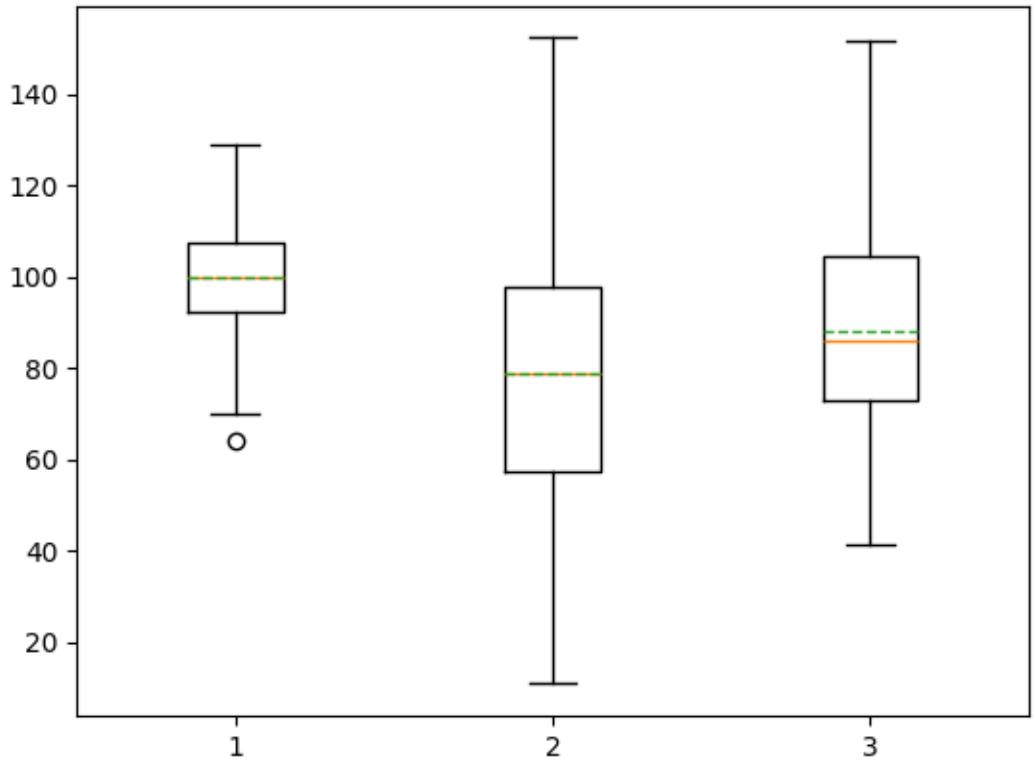

[5]:
```
plt.boxplot(values,vert=False)   #to consider all the outliers under the range
plt.show()
```




[6]:
```
plt.boxplot(values, meanline=True, showmeans=True, vert=True)   #to show the␣
 ↪mean of the datapoints
plt.show()
```




[7]:
```
#to plot multiple boxplots in one plane
import numpy as np
collectn_1 = np.random.normal(100, 10, 200)   #random generation of datapoints
collectn_2 = np.random.normal(80, 30, 200)
collectn_3 = np.random.normal(90, 20, 200)
values = [collectn_1, collectn_2, collectn_3]    #list of lists of datapoints
plt.boxplot(values,showmeans=True,meanline=True)
plt.show()
```




[8]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import statistics as st
df = pd.read_csv("train.csv")
df
```

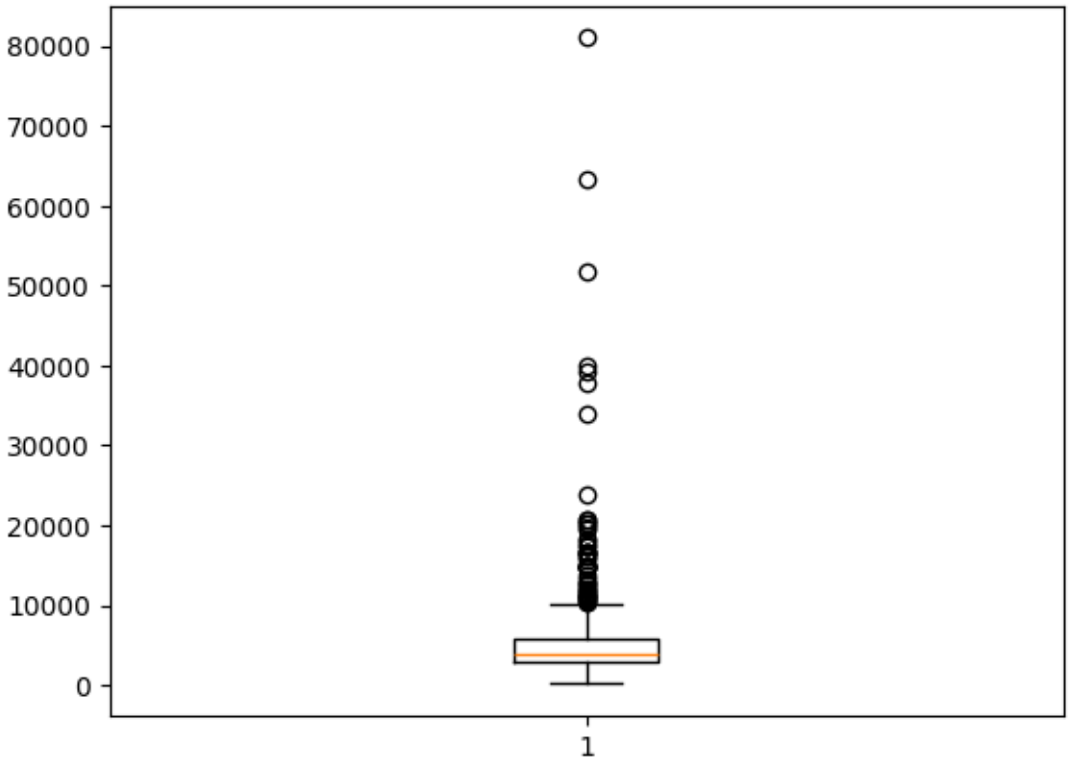
[8]:

| | Loan_ID | Gender | Married | Dependents | Education | Self_Employed | \ |
|---|---|---|---|---|---|---|---|
| 0 | LP001002 | Male | No | 0 | Graduate | No | |
| 1 | LP001003 | Male | Yes | 1 | Graduate | No | |
| 2 | LP001005 | Male | Yes | 0 | Graduate | Yes | |
| 3 | LP001006 | Male | Yes | 0 | Not Graduate | No | |
| 4 | LP001008 | Male | No | 0 | Graduate | No | |
| .. | … | … | … | … | … | … | |
| 609 | LP002978 | Female | No | 0 | Graduate | No | |
| 610 | LP002979 | Male | Yes | 3+ | Graduate | No | |
| 611 | LP002983 | Male | Yes | 1 | Graduate | No | |
| 612 | LP002984 | Male | Yes | 2 | Graduate | No | |
| 613 | LP002990 | Female | No | 0 | Graduate | Yes | |

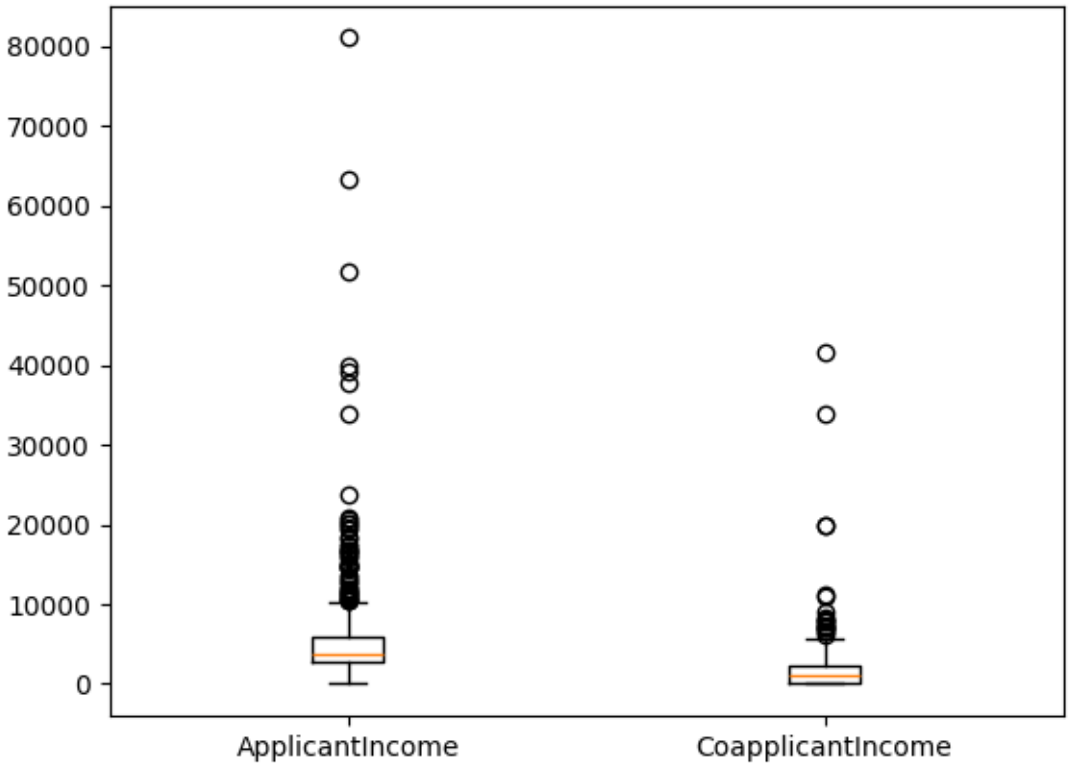
| | ApplicantIncome | CoapplicantIncome | LoanAmount | Loan_Amount_Term | \ |
|---|---|---|---|---|---|

```
0              5849                0.0         NaN             360.0
1              4583             1508.0       128.0             360.0
2              3000                0.0        66.0             360.0
3              2583             2358.0       120.0             360.0
4              6000                0.0       141.0             360.0
..              …                  …          …                …
609            2900                0.0        71.0             360.0
610            4106                0.0        40.0             180.0
611            8072              240.0       253.0             360.0
612            7583                0.0       187.0             360.0
613            4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              …           …         …
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

[9]:
```
values=df['ApplicantIncome']
plt.boxplot(values, vert=True)    #simple way to create a boxplot
plt.show()
```




[10]:
```
v1=df['ApplicantIncome']
v2=df['CoapplicantIncome']

values=[v1,v2]
plt.boxplot(values, vert=True,labels=['ApplicantIncome','CoapplicantIncome'])
plt.show()
```


80000
70000
60000
50000
40000
30000
20000
10000
0
ApplicantIncome
CoapplicantIncome