In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```python
dataset=pd.read_csv("iris.csv")
```

In [3]:
```python
dataset.head()
```

Out[3]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa

In [4]:
```python
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Sepal_Length  150 non-null    float64
 1   Sepal_Width   150 non-null    float64
 2   Petal_Length  150 non-null    float64
 3   Petal_Width   150 non-null    float64
 4   Species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

In [5]:
```python
dataset.describe()
```

Out[5]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

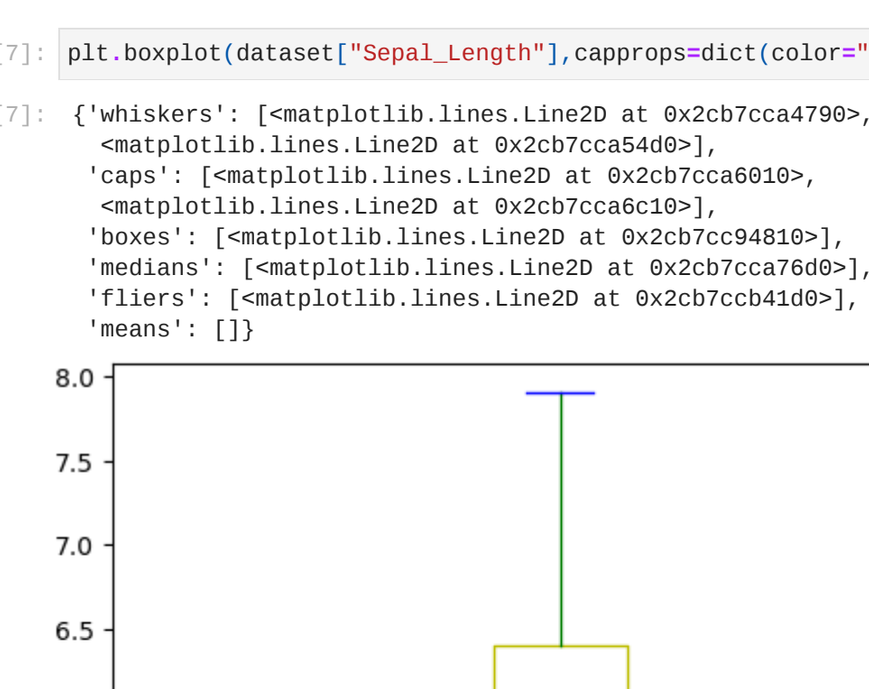
In [6]:
```python
dataset.isna().sum()
```

Out[6]:
```
Sepal_Length    0
Sepal_Width     0
Petal_Length    0
Petal_Width     0
Species         0
dtype: int64
```

As in the above cell we can see there is no null value in thge dataset so we don't have to worry about it

In [7]:
```python
plt.boxplot(dataset["Sepal_Length"],capprops=dict(color="b"),whiskerprops=dict(color="g"),boxprops=dict(color="y"),flierprops=dict(color="r"))
```

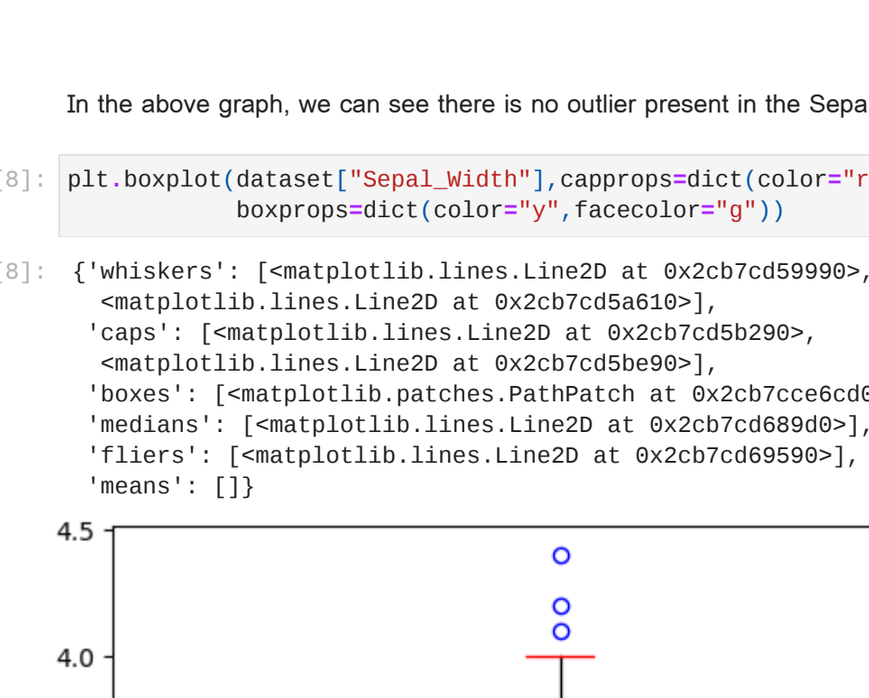
Out[7]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb7cca4790>,
  <matplotlib.lines.Line2D at 0x2cb7cca54d0>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb7cca6010>,
  <matplotlib.lines.Line2D at 0x2cb7cca6c10>],
 'boxes': [<matplotlib.lines.Line2D at 0x2cb7cc94810>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb7cca76d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb7ccb41d0>],
 'means': []}
```



In the above graph, we can see there is no outlier present in the Sepal_Length column in the dataset

In [8]:
```python
plt.boxplot(dataset["Sepal_Width"],capprops=dict(color="r"),flierprops=dict(color="y",markeredgecolor="b"),patch_artist=True,
            boxprops=dict(color="y",facecolor="g"))
```

Out[8]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb7cd59990>,
  <matplotlib.lines.Line2D at 0x2cb7cd5a610>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb7cd5b290>,
  <matplotlib.lines.Line2D at 0x2cb7cd5be90>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb7cce6cd0>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb7cd689d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb7cd69590>],
 'means': []}
```

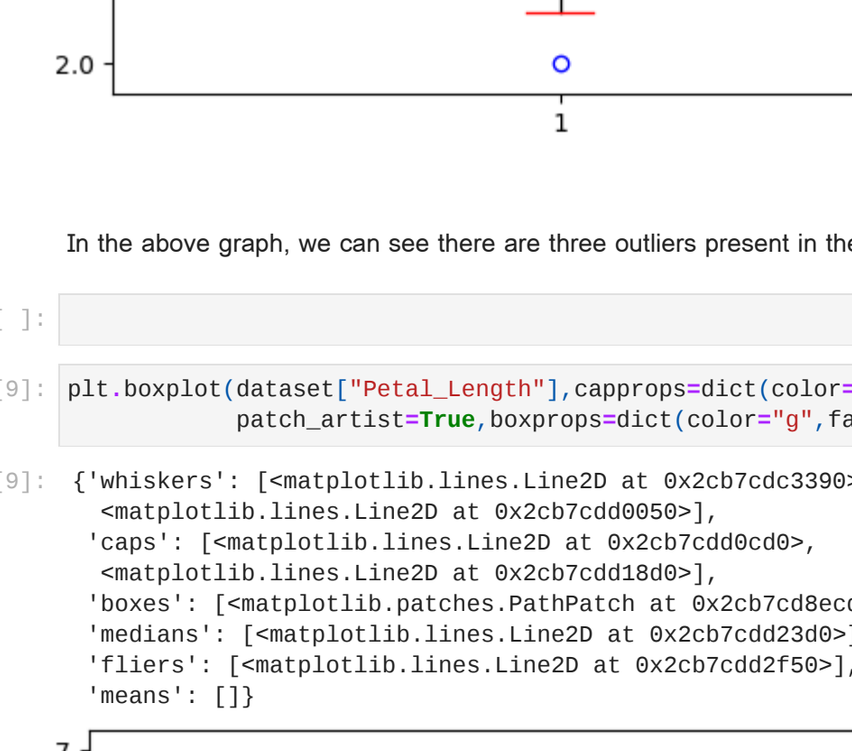


In the above graph, we can see there are three outliers present in the maximum side and one at the minimum side in Sepal_Width column in the dataset

In [ ]:

In [9]:
```python
plt.boxplot(dataset["Petal_Length"],capprops=dict(color="y"),flierprops=dict(color="b",markeredgecolor="r"),
            patch_artist=True,boxprops=dict(color="g",facecolor="y"),medianprops=dict(color="b"))
```

Out[9]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb7cdc3390>,
  <matplotlib.lines.Line2D at 0x2cb7cdd0050>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb7cdd0cd0>,
  <matplotlib.lines.Line2D at 0x2cb7cdd18d0>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb7cd8ecd0>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb7cdd23d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb7cdd2f50>],
 'means': []}
```

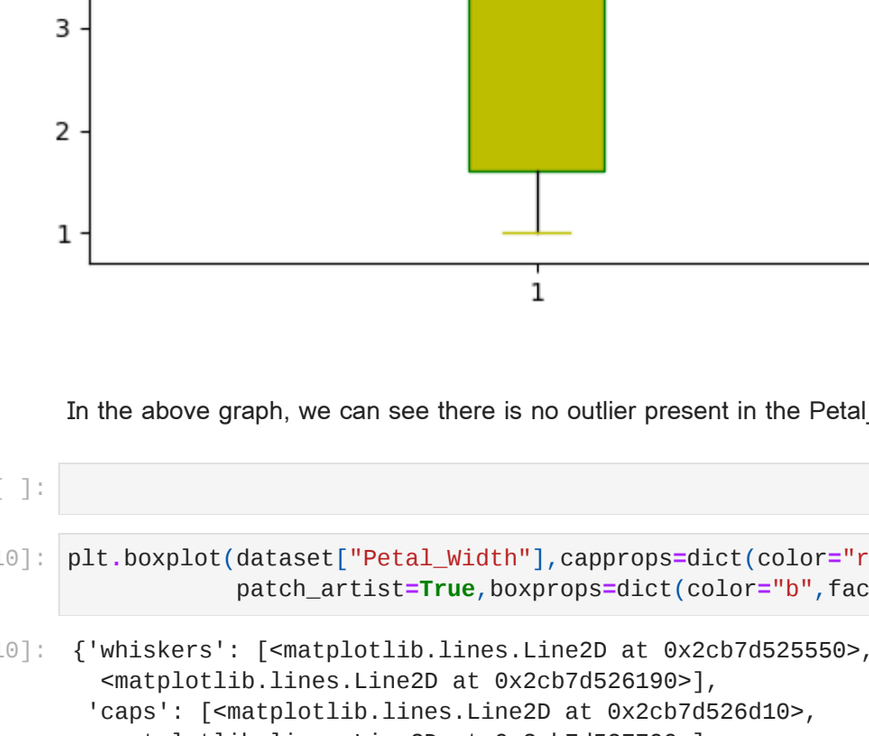


In the above graph, we can see there is no outlier present in the Petal_Length column in the dataset

In [ ]:

In [10]:
```python
plt.boxplot(dataset["Petal_Width"],capprops=dict(color="r"),flierprops=dict(color="b",markeredgecolor="g"),
            patch_artist=True,boxprops=dict(color="b",facecolor="g"))
```

Out[10]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb7d525550>,
  <matplotlib.lines.Line2D at 0x2cb7d526190>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb7d526d10>,
  <matplotlib.lines.Line2D at 0x2cb7d527790>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb7cde9610>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb7d52c250>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb7d52cd90>],
 'means': []}
```



In the above graph, we can see there is no outlier present in the Petal_width column in the dataset

In [ ]:

Cars93 dataset

In [13]:
```python
carsdata=pd.read_csv("cars93.csv")
```

In [14]:
```python
carsdata.head()
```

Out[14]:

	Manufacturer	Model	Type	Price	MPG.city	AirBags	Horsepower	Passengers	Rear.seat.room	Luggage.room
0	Acura	Integra	Small	15.9	25	Driver only	140	5	26.5	11.0
1	Acura	Legend	Midsize	33.9	18	Driver & Passenger	200	5	30.0	15.0
2	Audi	90	Compact	29.1	20	Driver only	172	5	28.0	14.0
3	Audi	100	Midsize	37.7	19	Driver & Passenger	172	6	31.0	17.0
4	BMW	535i	Midsize	30.0	22	Driver only	208	4	27.0	13.0

In [16]:
```python
carsdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 93 entries, 0 to 92
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Manufacturer    93 non-null     object 
 1   Model           93 non-null     object 
 2   Type            93 non-null     object 
 3   Price           93 non-null     float64
 4   MPG.city        93 non-null     int64  
 5   AirBags         79 non-null     object 
 6   Horsepower      93 non-null     int64  
 7   Passengers      93 non-null     int64  
 8   Rear.seat.room  91 non-null     float64
 9   Luggage.room    82 non-null     float64
dtypes: float64(3), int64(3), object(4)
memory usage: 7.4+ KB
```

In [17]:
```python
carsdata.isna().sum()
```

Out[17]:
```
Manufacturer       0
Model              0
Type               0
Price              0
MPG.city           0
AirBags           14
Horsepower         0
Passengers         0
Rear.seat.room     2
Luggage.room      11
dtype: int64
```

In [18]:
```python
carsdata.describe()
```

Out[18]:

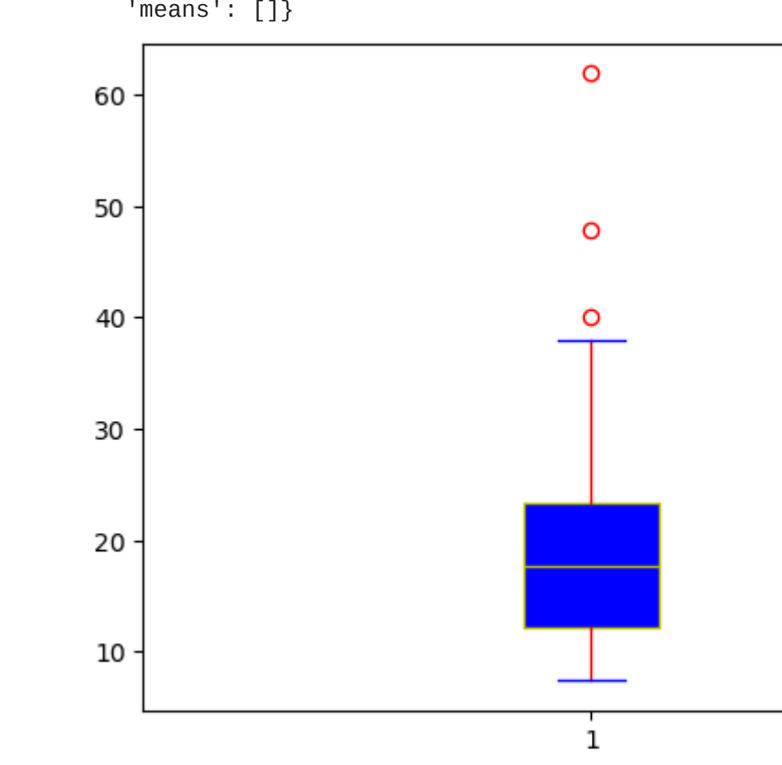
	Price	MPG.city	Horsepower	Passengers	Rear.seat.room	Luggage.room
count	93.000000	93.000000	93.000000	93.000000	91.000000	82.000000
mean	19.509677	22.365591	143.827957	5.086022	27.829670	13.890244
std	9.659430	5.619812	52.374410	1.038979	2.989072	2.997967
min	7.400000	15.000000	55.000000	2.000000	19.000000	6.000000
25%	12.200000	18.000000	103.000000	4.000000	26.000000	12.000000
50%	17.700000	21.000000	140.000000	5.000000	27.500000	14.000000
75%	23.300000	25.000000	170.000000	6.000000	30.000000	15.000000
max	61.900000	46.000000	300.000000	8.000000	36.000000	22.000000

In [19]:
```python
carsdata.shape
```

Out[19]:
```
(93, 10)
```

In [28]:
```python
plt.boxplot(carsdata["Price"],capprops=dict(color="b"),patch_artist=True,boxprops=dict(color="y",facecolor="b"),whiskerprops=dict(color="r"),
            flierprops=dict(color="g",markeredgecolor="r"),
            medianprops=dict(color="y"))
```

Out[28]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb07321a10>,
  <matplotlib.lines.Line2D at 0x2cb071c8f10>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb071cb590>,
  <matplotlib.lines.Line2D at 0x2cb07339590>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb07201610>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07339290>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb07339ed0>],
 'means': []}
```

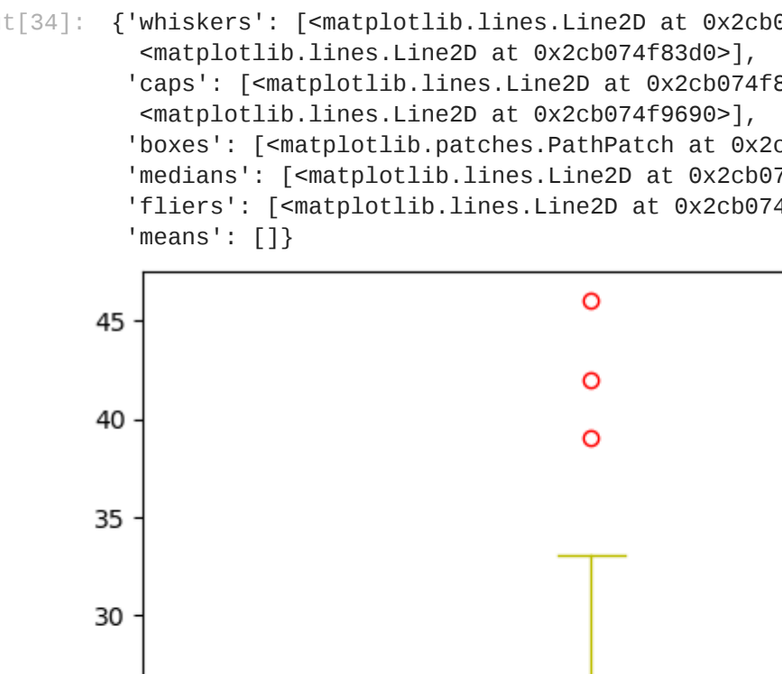


In the above graph, we can see there are three outliers present in the maximum side in "Price" column in the dataset

In [ ]:

In [34]:
```python
plt.boxplot(carsdata["MPG.city"],capprops=dict(color="y"),patch_artist=True,boxprops=dict(color="g",facecolor="r"),
            whiskerprops=dict(color="y"),
            flierprops=dict(color="b",markeredgecolor="r"),
            medianprops=dict(color="y"))
```

Out[34]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb074eba90>,
  <matplotlib.lines.Line2D at 0x2cb074f83d0>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb074f8d50>,
  <matplotlib.lines.Line2D at 0x2cb074f9690>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb0749d290>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb074f9f10>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb074fa750>],
 'means': []}
```

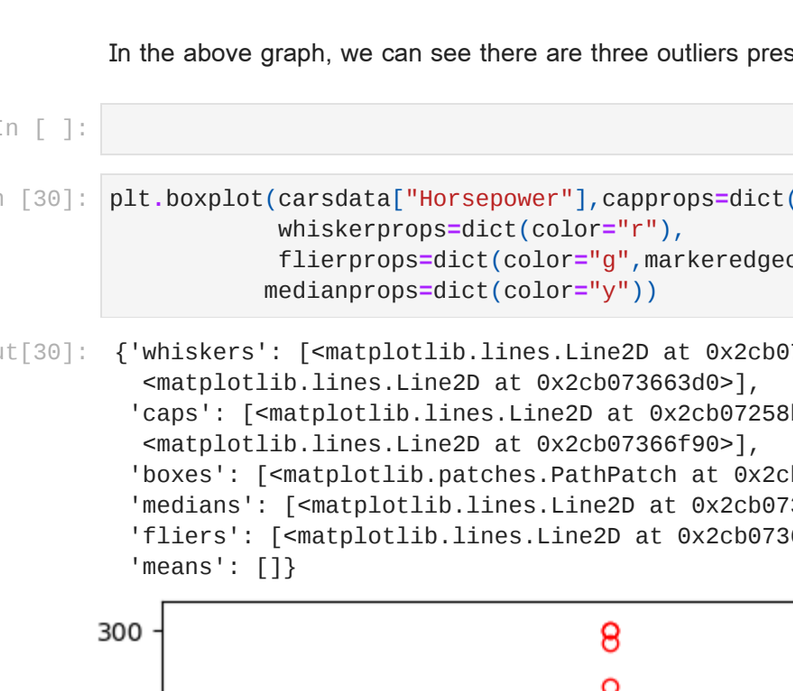


In the above graph, we can see there are three outliers present in the maximum side in "MPG.city" column in the dataset

In [ ]:

In [30]:
```python
plt.boxplot(carsdata["Horsepower"],capprops=dict(color="b"),patch_artist=True,boxprops=dict(color="y",facecolor="b"),
            whiskerprops=dict(color="r"),
            flierprops=dict(color="g",markeredgecolor="r"),
            medianprops=dict(color="y"))
```

Out[30]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb073659d0>,
  <matplotlib.lines.Line2D at 0x2cb073663d0>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb07258b10>,
  <matplotlib.lines.Line2D at 0x2cb07366f90>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb0727b250>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07367750>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb07367f50>],
 'means': []}
```

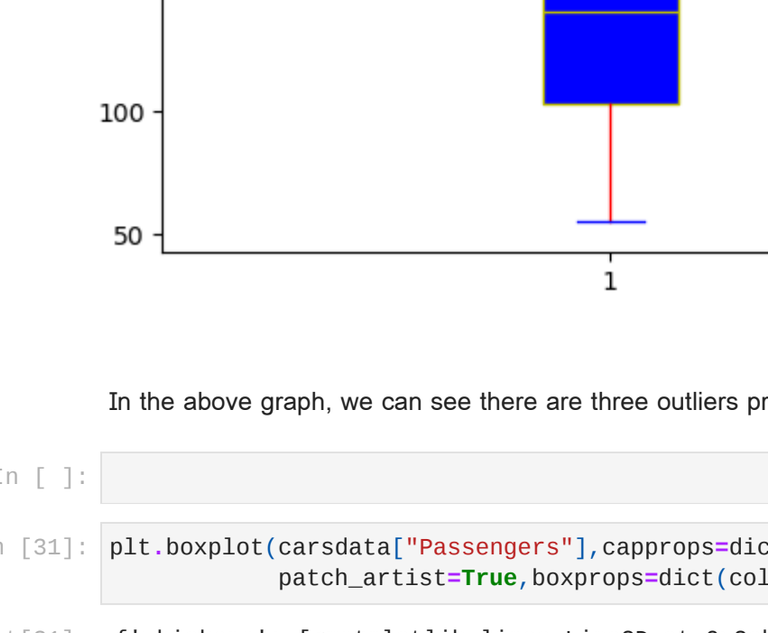


In the above graph, we can see there are three outliers present in the maximum side in "Horsepower" column in the dataset

In [ ]:

In [31]:
```python
plt.boxplot(carsdata["Passengers"],capprops=dict(color="r"),flierprops=dict(color="b",markeredgecolor="g"),
            patch_artist=True,boxprops=dict(color="b",facecolor="g"))
```

Out[31]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb07401ad0>,
  <matplotlib.lines.Line2D at 0x2cb07402310>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb07402c10>,
  <matplotlib.lines.Line2D at 0x2cb07403490>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb05fcb350>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07403cd0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb074105d0>],
 'means': []}
```

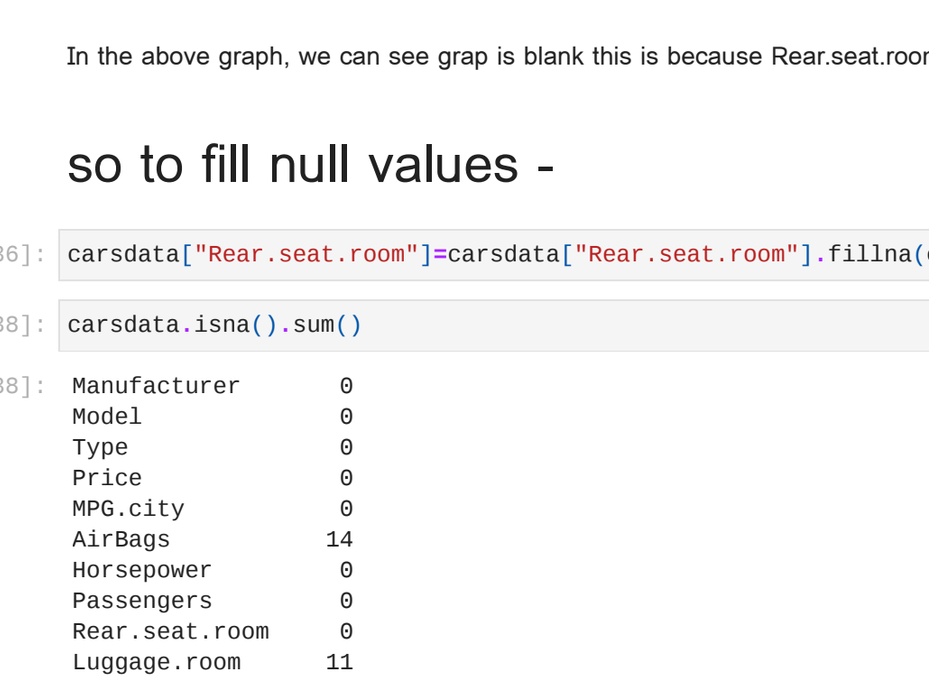


In the above graph, we can see there is no outlier present in the "Passengers" column in the dataset

In [ ]:

In [32]:
```python
plt.boxplot(carsdata["Rear.seat.room"],capprops=dict(color="g"),flierprops=dict(color="y",markeredgecolor="y"),
            patch_artist=True,boxprops=dict(color="r",facecolor="y"))
```

Out[32]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb073c89d0>,
  <matplotlib.lines.Line2D at 0x2cb073c9c50>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb073ce210>,
  <matplotlib.lines.Line2D at 0x2cb073cfed0>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb073e4e10>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07411810>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb073cf050>],
 'means': []}
```



In the above graph, we can see grap is blank this is because Rear.seat.room column has sum null values in the dataset

so to fill null values -

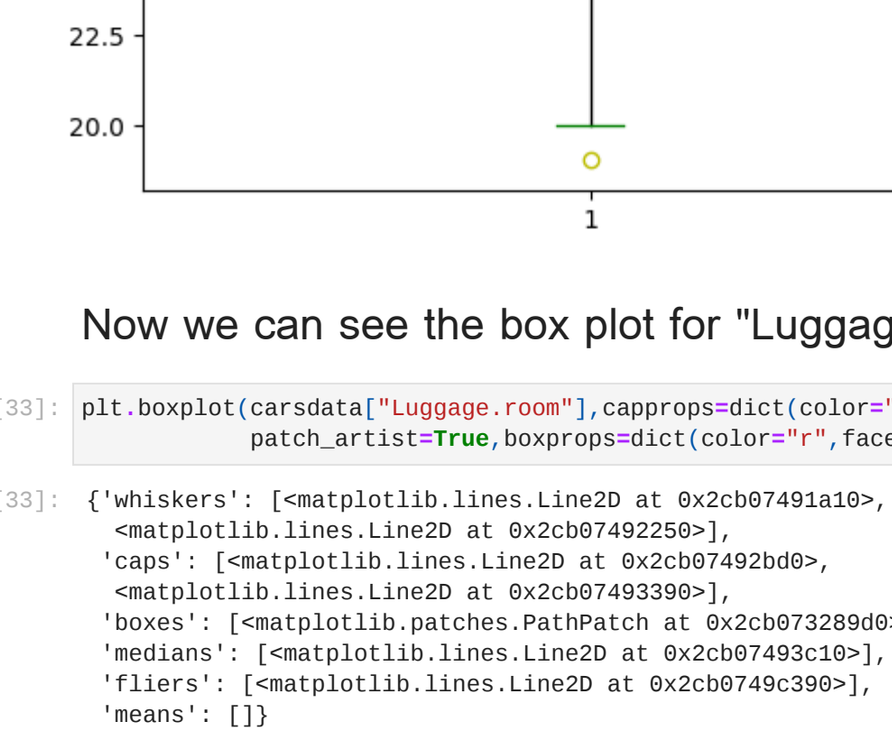
In [36]:
```python
carsdata["Rear.seat.room"]=carsdata["Rear.seat.room"].fillna(carsdata["Rear.seat.room"].mean())
```

In [38]:
```python
carsdata.isna().sum()
```

Out[38]:
```
Manufacturer       0
Model              0
Type               0
Price              0
MPG.city           0
AirBags           14
Horsepower         0
Passengers         0
Rear.seat.room     0
Luggage.room      11
dtype: int64
```

In [40]:
```python
plt.boxplot(carsdata["Rear.seat.room"],capprops=dict(color="g"),flierprops=dict(color="y",markeredgecolor="y"),
            patch_artist=True,boxprops=dict(color="r",facecolor="y"))
```

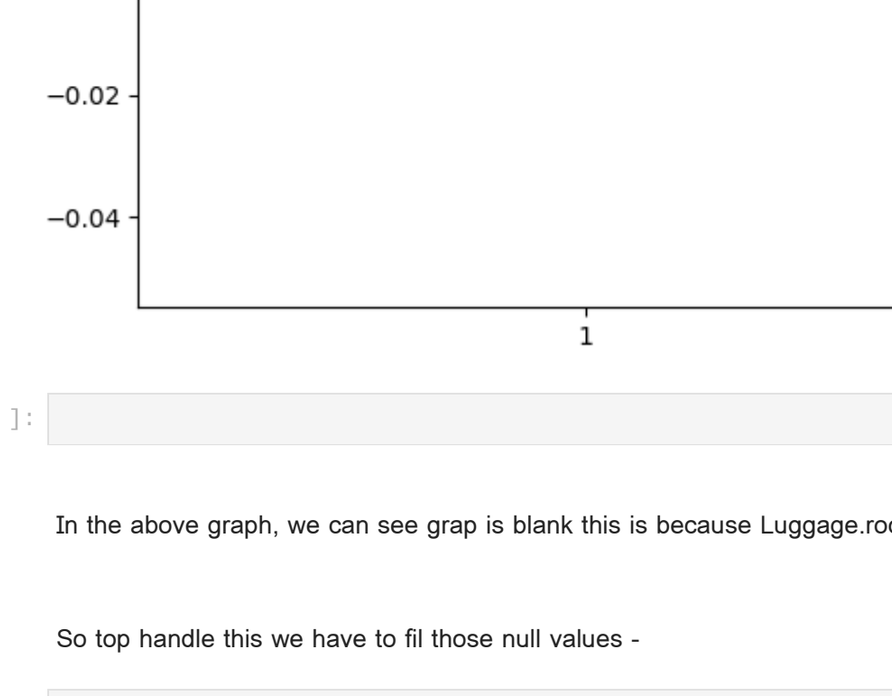
Out[40]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb07be52d0>,
  <matplotlib.lines.Line2D at 0x2cb07be6090>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb07be7450>,
  <matplotlib.lines.Line2D at 0x2cb07bf8690>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb07bce3d0>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07bf9510>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb07bfa3d0>],
 'means': []}
```



Now we can see the box plot for "Luggage.room" column in the dataset

In [33]:
```python
plt.boxplot(carsdata["Luggage.room"],capprops=dict(color="g"),flierprops=dict(color="y",markeredgecolor="y"),
            patch_artist=True,boxprops=dict(color="r",facecolor="y"))
```

Out[33]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb07491a10>,
  <matplotlib.lines.Line2D at 0x2cb07492250>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb07492bd0>,
  <matplotlib.lines.Line2D at 0x2cb07493390>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb073289d0>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07493c10>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb0749c390>],
 'means': []}
```



In [ ]:

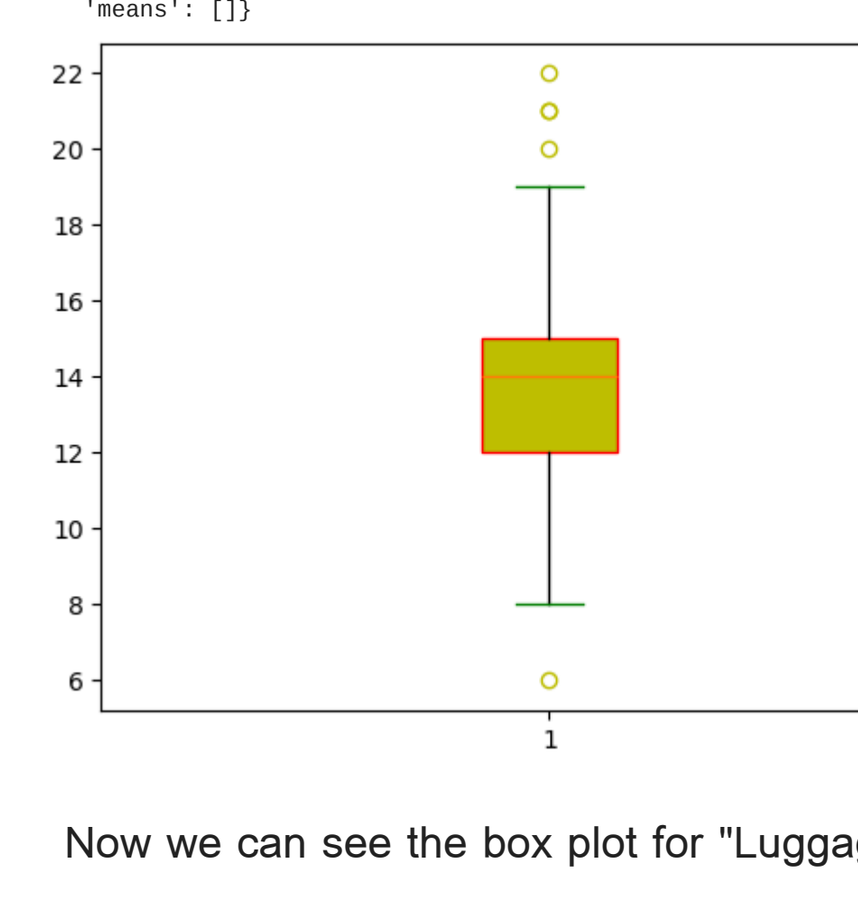
In the above graph, we can see grap is blank this is because Luggage.room column has sum null values in the dataset

So top handle this we have to fil those null values -

In [41]:
```python
carsdata["Luggage.room"]=carsdata["Luggage.room"].fillna(carsdata["Luggage.room"].mean())
```

In [42]:
```python
plt.boxplot(carsdata["Luggage.room"],capprops=dict(color="g"),flierprops=dict(color="y",markeredgecolor="y"),
            patch_artist=True,boxprops=dict(color="r",facecolor="y"))
```

Out[42]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2cb07c64090>,
  <matplotlib.lines.Line2D at 0x2cb07c65110>],
 'caps': [<matplotlib.lines.Line2D at 0x2cb07c65f90>,
  <matplotlib.lines.Line2D at 0x2cb07c66c90>],
 'boxes': [<matplotlib.patches.PathPatch at 0x2cb07c59f10>],
 'medians': [<matplotlib.lines.Line2D at 0x2cb07c67a90>],
 'fliers': [<matplotlib.lines.Line2D at 0x2cb07c68a10>],
 'means': []}
```



Now we can see the box plot for "Luggage.room" column in the dataset

Parameters used in this box plot visualization are ass follows-

1)capprops-It is related to minimum and maximum value data

2)whiskeprops=It is related to the verticle line between minimum and 1st quartile, and 3rd quartile and the maximum in the data

3)boxprops=It is related to the whisker(main box) in the data

4)flierprops=It is related to the Outliers that are present in the data