In [6]:
```
cd datascience_internship
```

```
G:\datascience_internship
```

In [7]:
```
import pandas as pd
import numpy as np
```

In [72]:
```
df = pd.read_csv("BBC.csv")
df.head(10)
```

Out[72]:

	MaritalStatus	Gender	YearlyIncome	TotalChildren	NumberChildrenAtHome	EnglishEducation	H
0	5	1.0	90000	2	0	5	
1	5	1.0	60000	3	3	5	
2	5	1.0	60000	3	3	5	
3	5	NaN	70000	0	0	5	
4	5	2.0	80000	5	5	5	
5	5	1.0	70000	0	0	5	
6	5	2.0	70000	0	0	5	
7	5	1.0	60000	3	3	5	
8	5	2.0	60000	4	4	5	
9	5	1.0	70000	0	0	5	

In [43]:
```
df.isna().sum()
```

Out[43]:
```
MaritalStatus           0
Gender                  0
YearlyIncome            0
TotalChildren           0
NumberChildrenAtHome    0
EnglishEducation        0
HouseOwnerFlag          0
NumberCarsOwned         0
CommuteDistance         0
Region                  0
Age                     0
BikeBuyer               0
dtype: int64
```

In [44]:
```
df["Gender"][3]=1
```

```
C:\Users\Admin\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWith
CopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  """Entry point for launching an IPython kernel.
```

In [45]:
```
df.dtypes
```

Out[45]:
```
MaritalStatus            int64
Gender                 float64
YearlyIncome             int64
TotalChildren            int64
NumberChildrenAtHome     int64
EnglishEducation         int64
HouseOwnerFlag           int64
NumberCarsOwned          int64
CommuteDistance          int64
Region                   int64
Age                      int64
BikeBuyer                int64
dtype: object
```

In [47]:
```
df["Gender"].value_counts()
```

Out[47]:
```
1.0    9352
2.0    9132
Name: Gender, dtype: int64
```

In [49]:
```
df["YearlyIncome"].value_counts()
```

Out[49]:
```
60000     3127
40000     2747
70000     2349
30000     2287
20000     1767
80000     1342
10000     1155
90000      842
50000      670
100000     571
130000     512
110000     474
120000     332
170000     112
150000     103
160000      94
Name: YearlyIncome, dtype: int64
```

In [50]:
```
df["YearlyIncome"].mean()
```

Out[50]:
```
57305.77797013633
```

In [56]:
```
df["BikeBuyer"].value_counts()
```

Out[56]:
```
0    9352
1    9132
Name: BikeBuyer, dtype: int64
```

In [60]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==160000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
58
```

In [63]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==70000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
1290
```

In [61]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==60000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
1578
```

In [62]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==40000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
1467
```

In [64]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==30000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
1152
```

In [65]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==20000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
748
```

In [66]:
```
c=0
for i in range(len(df)):
    if df["YearlyIncome"][i]==10000:
        #print(df["BikeBuyer"][i])
        if df["BikeBuyer"][i] == 1:
            c = c+1
print(c)
```

```
438
```

In [67]:
```
df["TotalChildren"].unique()
```

Out[67]:
```
array([2, 3, 0, 5, 4, 1], dtype=int64)
```

In [68]:
```
df["NumberChildrenAtHome"].unique()
```

Out[68]:
```
array([0, 3, 5, 4, 1, 2], dtype=int64)
```

In [69]:
```
df["TotalChildren"].value_counts()
```

Out[69]:
```
0    5165
2    3779
1    3619
4    2303
3    2194
5    1424
Name: TotalChildren, dtype: int64
```

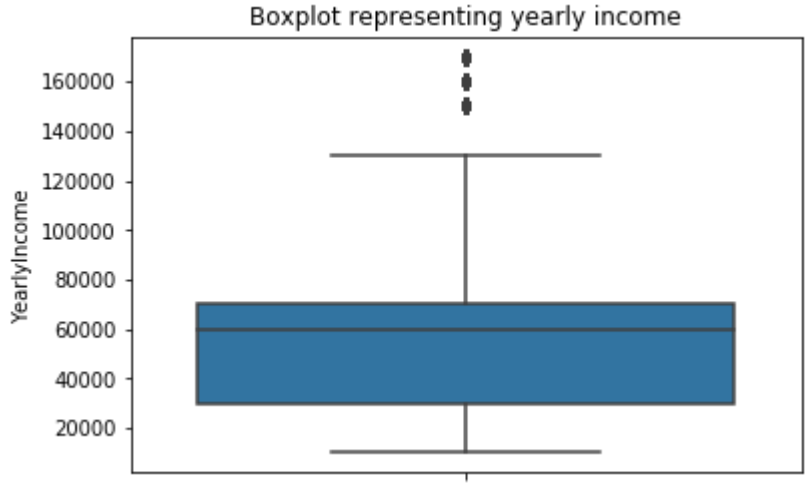
In [70]:
```
df["NumberChildrenAtHome"].value_counts()
```

Out[70]:
```
0    11116
1     2460
2     1648
3     1204
4     1089
5      967
Name: NumberChildrenAtHome, dtype: int64
```

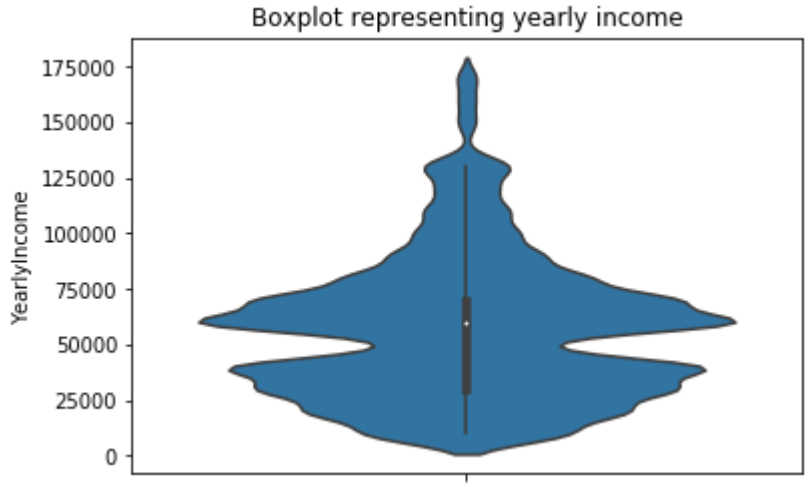
In [71]:
```
print(df["TotalChildren"],df["NumberChildrenAtHome"])
```

```
0        2
1        3
2        3
3        0
4        5
        ..
18479    1
18480    3
18481    3
18482    3
18483    0
Name: TotalChildren, Length: 18484, dtype: int64 0        0
1        3
2        3
3        0
4        5
        ..
18479    0
18480    0
18481    0
18482    0
18483    0
Name: NumberChildrenAtHome, Length: 18484, dtype: int64
```

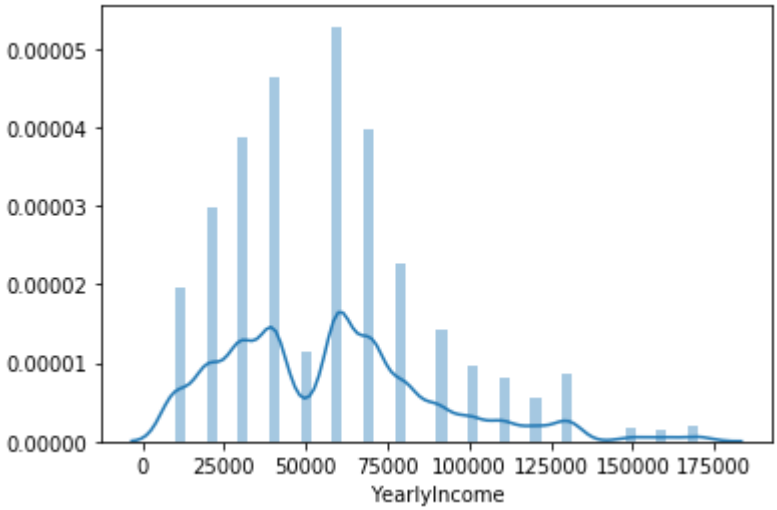
In [74]:
```
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.boxplot(y=df["YearlyIncome"])
plt.title("Boxplot representing yearly income")
plt.show()
```



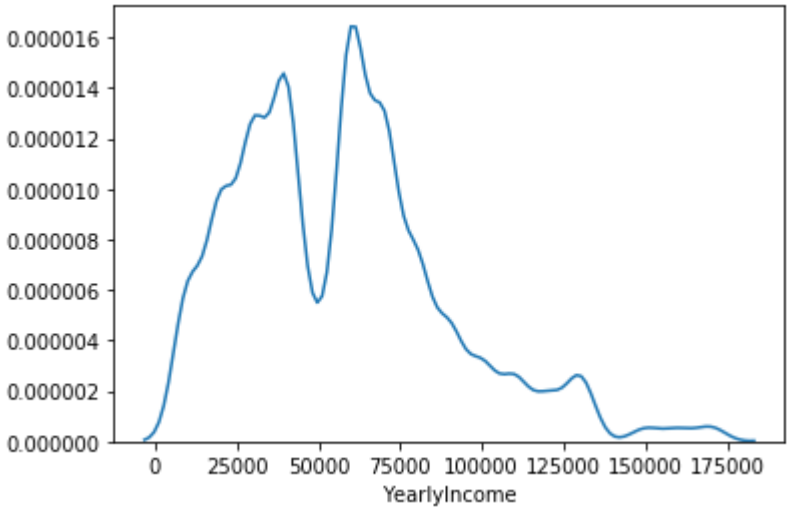
In [75]:
```
sns.violinplot(y=df["YearlyIncome"])
plt.title("Boxplot representing yearly income")
plt.show()
```



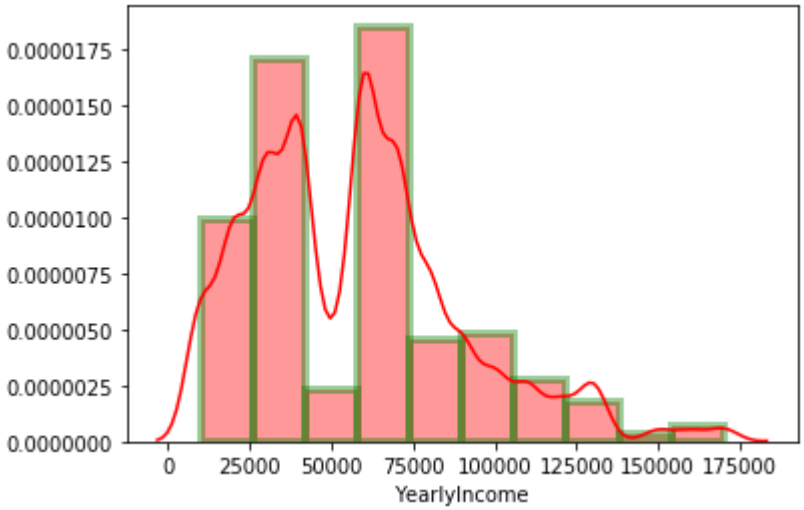
In [76]:
```
sns.distplot(df["YearlyIncome"])
plt.show()
```



In [77]:
```
sns.distplot(df["YearlyIncome"],hist=False)
plt.show()
```



In [82]:
```
sns.distplot(df["YearlyIncome"],bins=10,color='r',hist_kws=dict(edgecolor="g",lir
plt.show()
```



In [ ]:

In [ ]: