In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns                    #visulization
import matplotlib.pyplot as plt          #visulization
sns.set(color_codes=True)
```

In [5]:

```
df=sns.load_dataset('tips')
df
```

Out[5]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

In [6]:

```
df.dtypes
```

Out[6]:

```
total_bill     float64
tip            float64
sex           category
smoker        category
day           category
time          category
size             int64
dtype: object
```

In [7]:
```
df=df.drop(['time'],axis=1)
df.head(5)
```

Out[7]:

	total_bill	tip	sex	smoker	day	size
0	16.99	1.01	Female	No	Sun	2
1	10.34	1.66	Male	No	Sun	3
2	21.01	3.50	Male	No	Sun	3
3	23.68	3.31	Male	No	Sun	2
4	24.59	3.61	Female	No	Sun	4

In [8]:
```
df.columns
```

Out[8]:
```
Index(['total_bill', 'tip', 'sex', 'smoker', 'day', 'size'], dtype='object')
```

In [9]:
```
df=df.rename(columns={'total_bill':'HP', 'tip':'EG', 'sex':'SE', 'smoker':'SMO', 'day':'DA', 'size':'SI'}
df.head(5)
```

Out[9]:

	HP	EG	SE	SMO	DA	SI
0	16.99	1.01	Female	No	Sun	2
1	10.34	1.66	Male	No	Sun	3
2	21.01	3.50	Male	No	Sun	3
3	23.68	3.31	Male	No	Sun	2
4	24.59	3.61	Female	No	Sun	4

In [10]:
```
df.shape
```

Out[10]:
```
(244, 6)
```

In [14]:
```
duplicate_rows=df[df.duplicated()]
print('numb of duplicate rows',duplicate_rows.shape)

```

```
numb of duplicate rows (1, 6)
```

In [16]:
```
df.count()
```

Out[16]:
```
HP      244
EG      244
SE      244
SMO     244
DA      244
SI      244
dtype: int64
```

In [17]:
```
df=df.drop_duplicates()
df.head()
```

Out[17]:

	HP	EG	SE	SMO	DA	SI
0	16.99	1.01	Female	No	Sun	2
1	10.34	1.66	Male	No	Sun	3
2	21.01	3.50	Male	No	Sun	3
3	23.68	3.31	Male	No	Sun	2
4	24.59	3.61	Female	No	Sun	4

In [18]:
```
df.count()
```

Out[18]:
```
HP      243
EG      243
SE      243
SMO     243
DA      243
SI      243
dtype: int64
```

In [19]:
```
df.isnull().sum()
```

Out[19]:
```
HP      0
EG      0
SE      0
SMO     0
DA      0
SI      0
dtype: int64
```

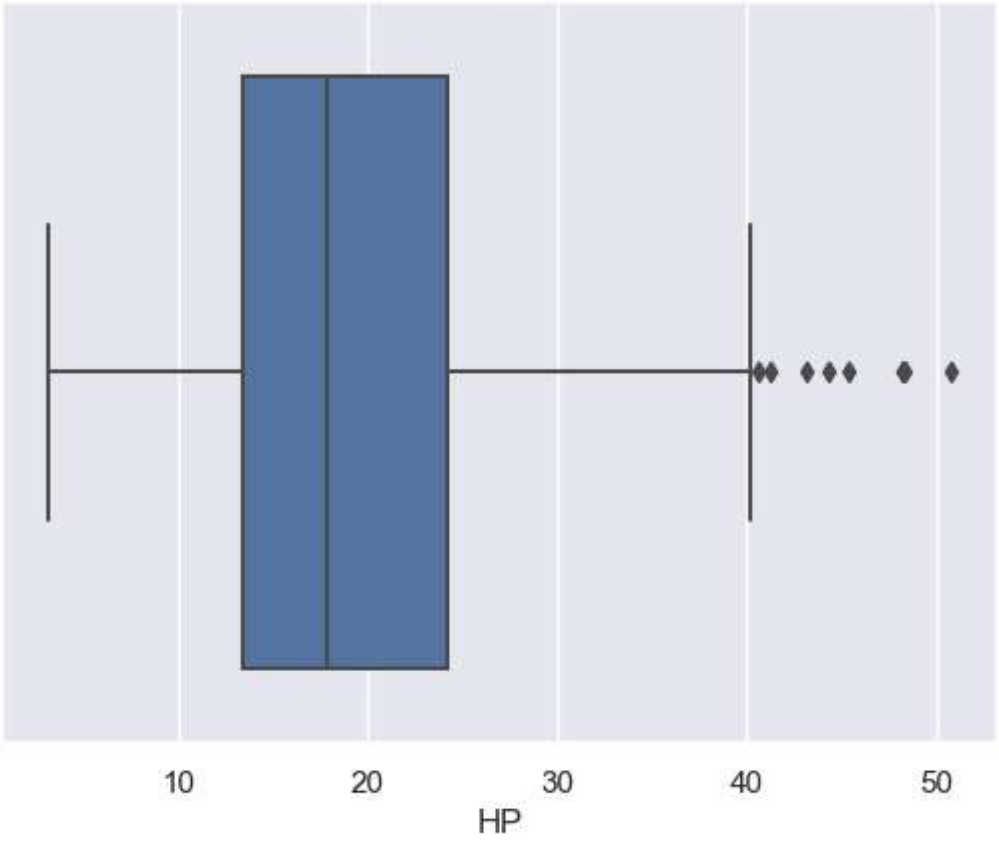
if rows null

In [20]:
```
df.dropna()
df.count()
```

Out[20]:
```
HP      243
EG      243
SE      243
SMO     243
DA      243
SI      243
dtype: int64
```

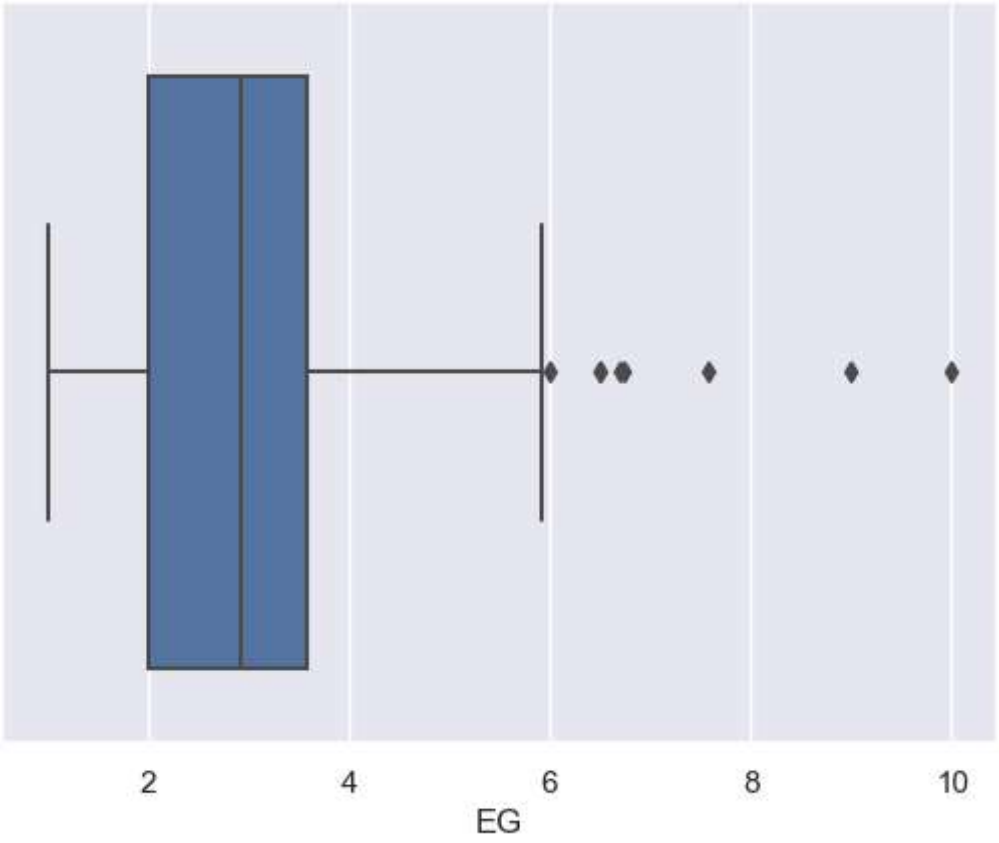
In [21]:
```
1 sns.boxplot(x=df['HP'])
```

Out[21]: <Axes: xlabel='HP'>



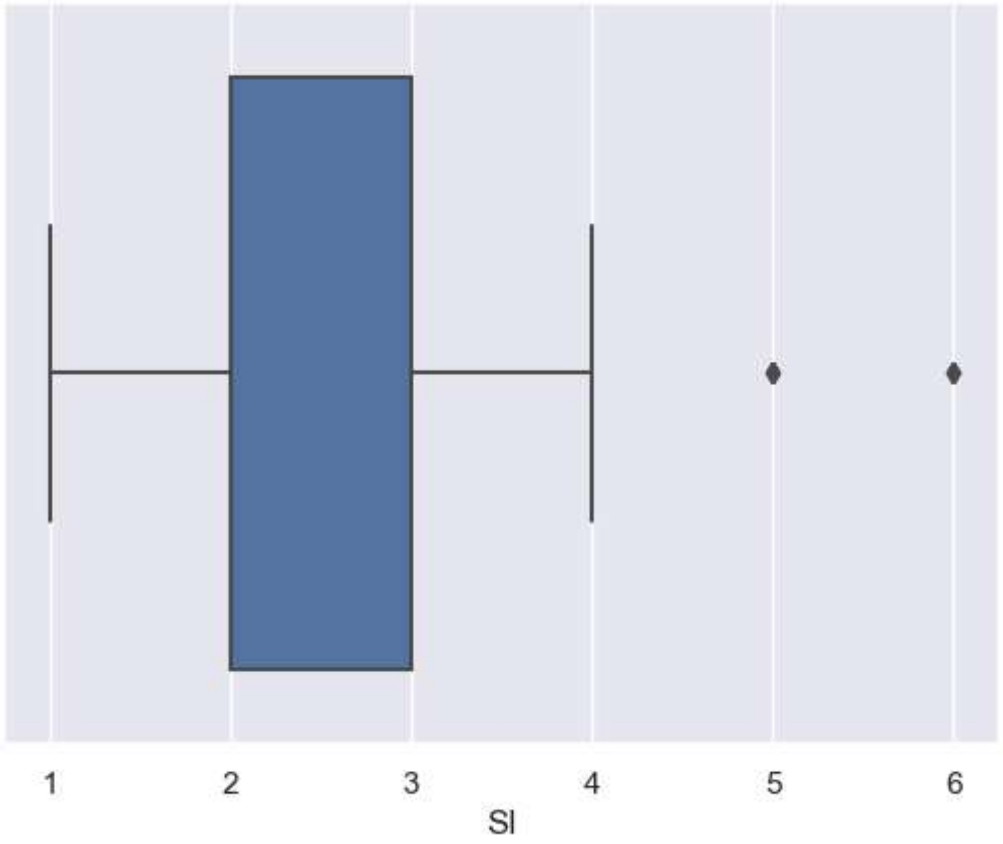
In [22]:
```
1 sns.boxplot(x=df['EG'])
```

Out[22]: <Axes: xlabel='EG'>



In [23]:
```
1 sns.boxplot(x=df['SI'])
```

Out[23]: <Axes: xlabel='SI'>



In [25]:
```
Q1=df.quantile(0.5)
Q3=df.quantile(0.75)
IQR=Q3-Q1
print(IQR)
```

```
HP    6.365
EG    0.655
SI    1.000
dtype: float64

C:\Users\Super\AppData\Local\Temp\ipykernel_10012\3515705857.py:1: FutureWarning: The default value of numer
ic_only in DataFrame.quantile is deprecated. In a future version, it will default to False. Select only vali
d columns or specify the value of numeric_only to silence this warning.
  Q1=df.quantile(0.5)
C:\Users\Super\AppData\Local\Temp\ipykernel_10012\3515705857.py:2: FutureWarning: The default value of numer
ic_only in DataFrame.quantile is deprecated. In a future version, it will default to False. Select only vali
d columns or specify the value of numeric_only to silence this warning.
  Q3=df.quantile(0.75)
```

Correlation of different features

In [26]:
```
df.corr()
```

```
C:\Users\Super\AppData\Local\Temp\ipykernel_10012\1134722465.py:1: FutureWarning: The default value of numer
ic_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid co
lumns or specify the value of numeric_only to silence this warning.
  df.corr()
```

Out[26]:

	HP	EG	SI
HP	1.000000	0.674998	0.597589
EG	0.674998	1.000000	0.488400
SI	0.597589	0.488400	1.000000

In [ ]:
```

```