In [ ]:

In [1]:
```
import numpy as np
import pandas as pd
```

In [3]:
```
df=pd.read_csv('C:/satish (coding)/csv files/train (1).csv')
```

In [4]:
```
df.head()
```

Out[4]:

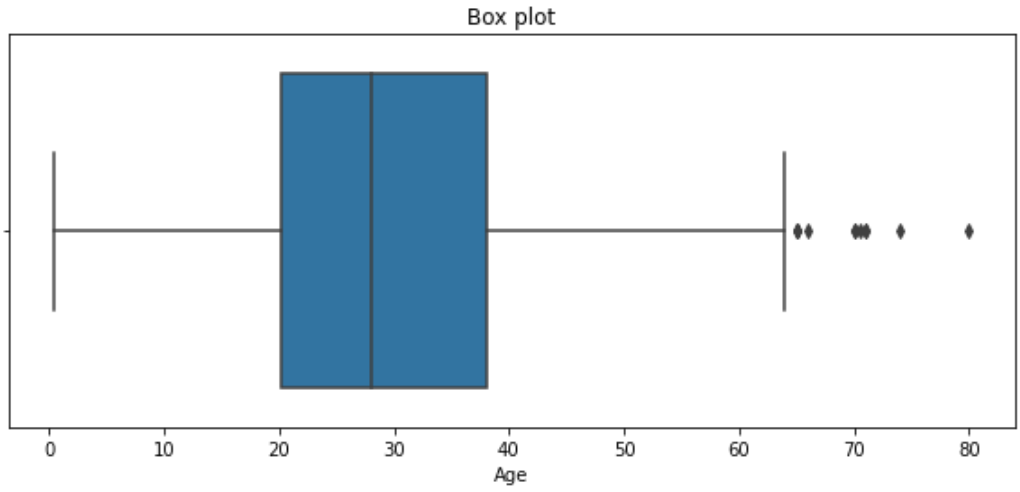
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [9]:
```
#IQR method

out = []
def iqr_outlier(df):
    q1 = df.quantile(0.25)
    q3 = df.quantile(0.75)
    iqr = q3-q1
    Lower_tail = q1-1.5*iqr
    Upper_tail = q3+1.5*iqr
    for i in df:
        if i> Upper_tail or i< Lower_tail:
            out.append(i)
    print("Outliers:",out)
iqr_outlier(df['Age'])
```

```
Outliers: [66.0, 65.0, 71.0, 70.5, 65.0, 65.0, 71.0, 80.0, 70.0, 70.0, 74.0]
```

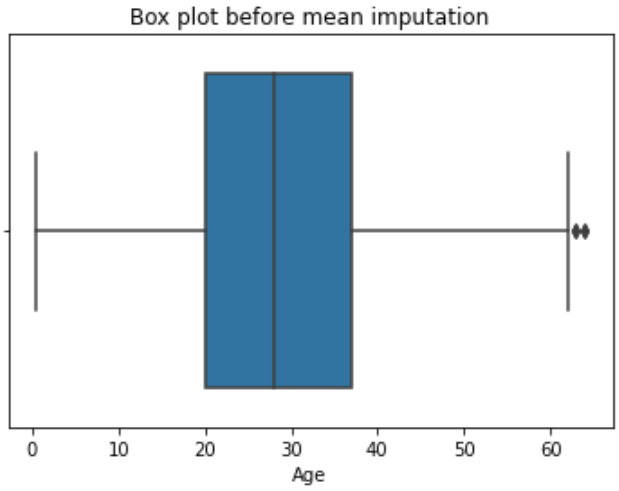
In [11]:
```
#Plotting Boxplot

import seaborn as sns
from matplotlib import pyplot as plt
from statsmodels.graphics.gofplots import qqplot

def Box_plots(df):
    plt.figure(figsize=(10,4))
    plt.title("Box plot")
    sns.boxplot(df)
    plt.show()
Box_plots(df['Age'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional a
rgument will be `data`, and passing other arguments without an explicit keyword will resul
t in an error or misinterpretation.
  warnings.warn(
```



In [18]:
```
# Mean imputation
sns.boxplot(df['Age'])
plt.title("Box plot before mean imputation")
plt.show()
q1 = df['Age'].quantile(0.25)
q3 = df['Age'].quantile(0.75)
iqr = q3-q1
Lower_tail = q1-1.5*iqr
Upper_tail = q3+1.5*iqr
m = np.mean(df['Age'])
for i in df['Age']:
    if i<Lower_tail or i>Upper_tail :
        df['Age'] = df['Age'].replace(i,m)
sns.boxplot(df['Age'])
plt.title("Box plot after mean imputation")
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional a
rgument will be `data`, and passing other arguments without an explicit keyword will resul
t in an error or misinterpretation.
  warnings.warn(
```




```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional a
rgument will be `data`, and passing other arguments without an explicit keyword will resul
t in an error or misinterpretation.
  warnings.warn(
```