In [2]:
```
import pandas as pd
#import data
#read_csv function is used to read test.csv file
data=pd.read_csv('test.csv')
```

In [3]:
```
data
```

Out[3]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	
...	...	...	...	...	...	...	...	...	...	...	
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	

418 rows × 11 columns

In [18]:
```
data=data[['Age','Fare','SibSp']] #data consists of 3 columns Age,Fare,SibSp
data.head()
```

Out[18]:

	Age	Fare	SibSp
0	34.5	7.8292	0
1	47.0	7.0000	1
2	62.0	9.6875	0
3	27.0	8.6625	0
4	22.0	12.2875	1

In [19]:
```
data.describe() #to see the five number summary we use da describe function
```

Out[19]:

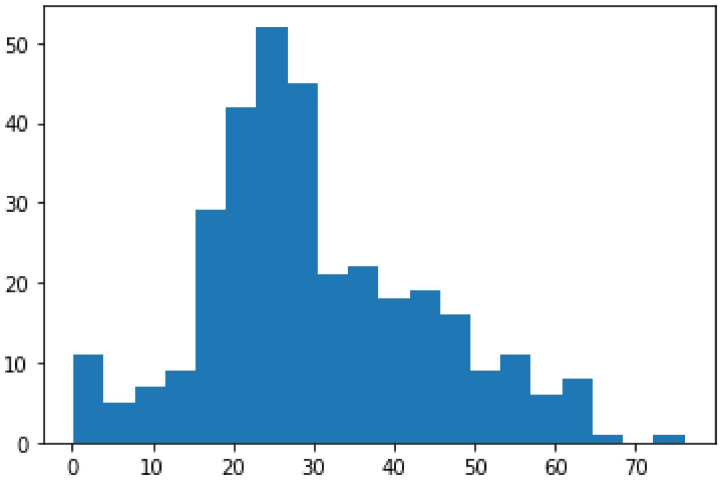
	Age	Fare	SibSp
count	332.000000	417.000000	418.000000
mean	30.272590	35.627188	0.447368
std	14.181209	55.907576	0.896760
min	0.170000	0.000000	0.000000
25%	21.000000	7.895800	0.000000
50%	27.000000	14.454200	0.000000
75%	39.000000	31.500000	1.000000
max	76.000000	512.329200	8.000000

In [20]:
```
data.skew()
```

Out[20]:
```
Age      0.457361
Fare     3.687213
SibSp    4.168337
dtype: float64
```

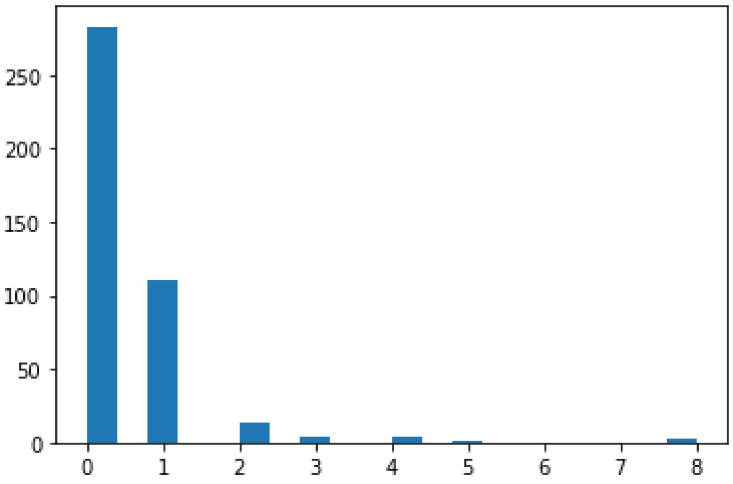
In [21]:
```
import matplotlib.pyplot as plt
plt.hist(data.Age,bins=20)
```

Out[21]:
```
(array([11.,  5.,  7.,  9., 29., 42., 52., 45., 21., 22., 18., 19., 16.,
         9., 11.,  6.,  8.,  1.,  0.,  1.]),
 array([ 0.17  ,  3.9615,  7.753 , 11.5445, 15.336 , 19.1275, 22.919 ,
        26.7105, 30.502 , 34.2935, 38.085 , 41.8765, 45.668 , 49.4595,
        53.251 , 57.0425, 60.834 , 64.6255, 68.417 , 72.2085, 76.    ]),
 <BarContainer object of 20 artists>)
```



In [22]:
```
plt.hist(data.SibSp,bins=20)
```

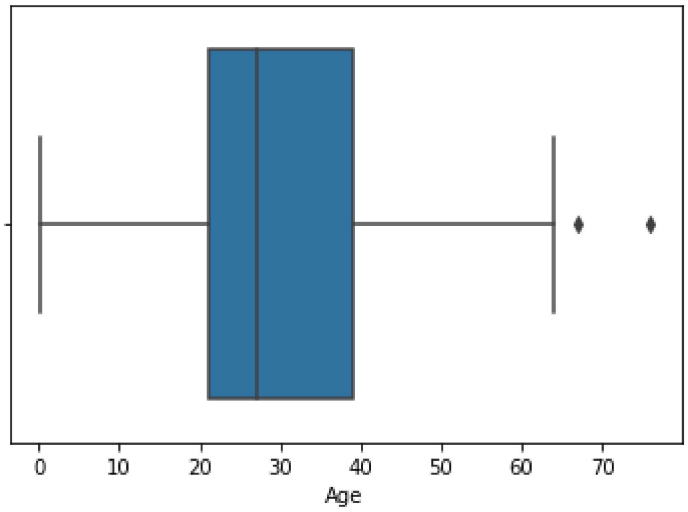
Out[22]:
```
(array([283.,   0., 110.,   0.,   0.,  14.,   0.,   4.,   0.,   0.,   4.,
          0.,   1.,   0.,   0.,   0.,   0.,   0.,   0.,   2.]),
 array([0. , 0.4, 0.8, 1.2, 1.6, 2. , 2.4, 2.8, 3.2, 3.6, 4. , 4.4, 4.8,
        5.2, 5.6, 6. , 6.4, 6.8, 7.2, 7.6, 8. ]),
 <BarContainer object of 20 artists>)
```



In [11]:
```
import seaborn as sns
sns.boxplot(data.Age)
```

```
C:\Users\sankr\Pictures\New folder\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 argument will be `data`, and passing other argume
nts without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

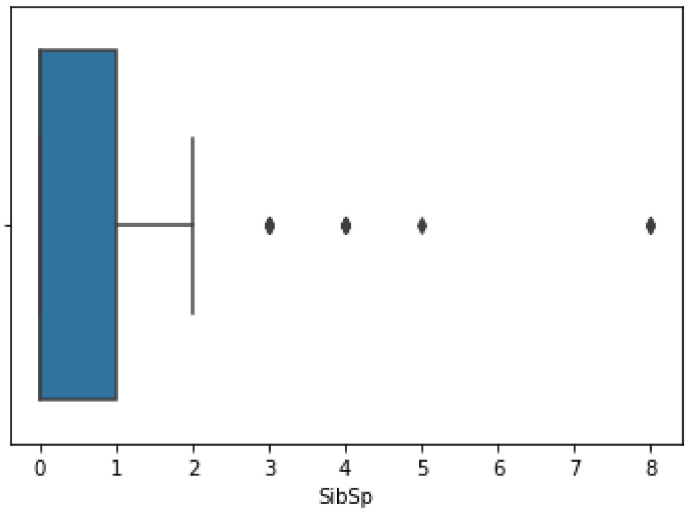
Out[11]: <AxesSubplot:xlabel='Age'>



In [23]:
```
sns.boxplot(data.SibSp)
```

```
C:\Users\sankr\Pictures\New folder\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 argument will be `data`, and passing other argume
nts without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[23]: <AxesSubplot:xlabel='SibSp'>



In [ ]: