In [1]:
```
import pandas as pd
df=pd.read_csv('Titanic-Dataset.csv')
```

In [2]:
```
df.head()
```

Out[2]:

	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 [3]:
```
df.tail()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

In [4]:
```
df.shape
```

Out[4]:
```
(891, 12)
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [6]:
```
df.describe()
```

Out[6]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

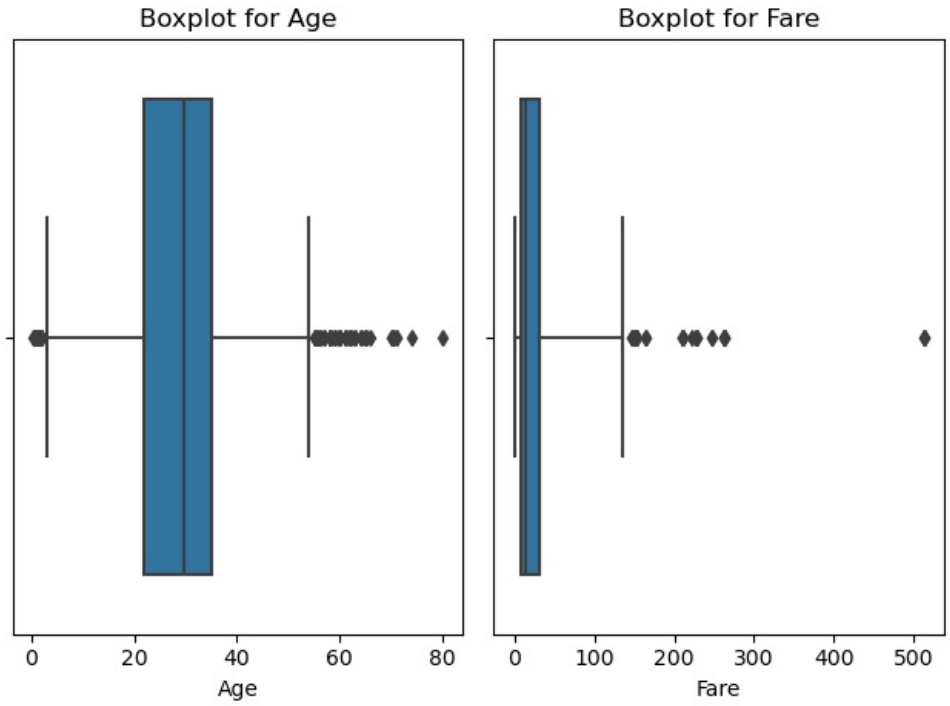
In [8]:
```
df['Age']=df['Age'].fillna(value=df['Age'].mean())
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [10]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [28]:
```
plt.subplot(1,2,1)
plt.title("Boxplot for Age")
sns.boxplot(data=df,x='Age')
plt.subplot(1,2,2)
plt.title("Boxplot for Fare")
sns.boxplot(data=df,x="Fare")
#for having space between the subplots
plt.tight_layout()
plt.show()
```

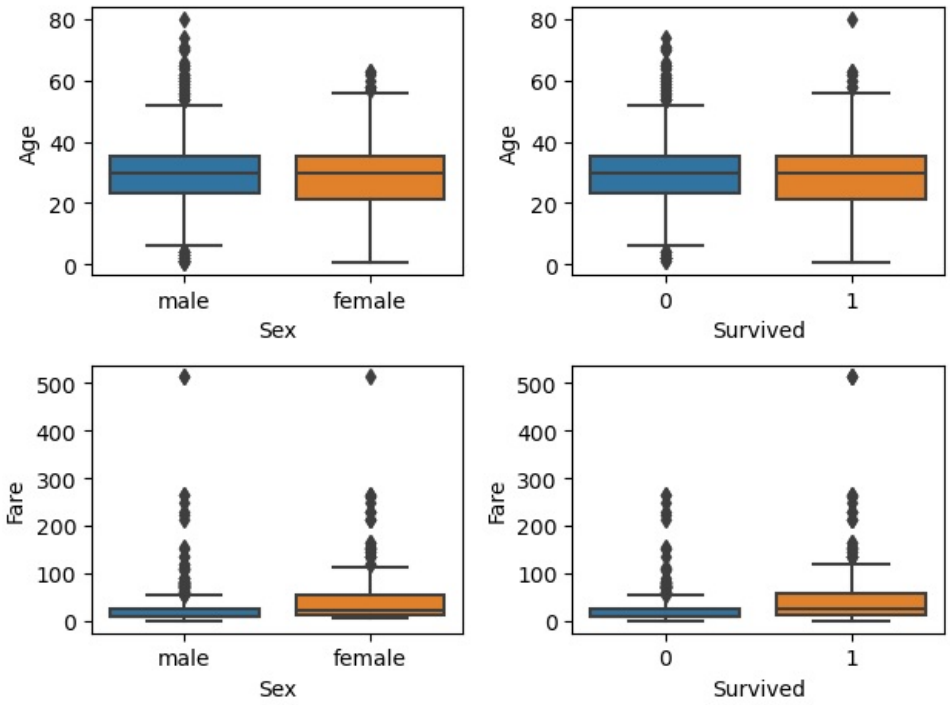


In [34]:
```
print("Conculsion for above plots are : ")
print("1) Almost all people belongs to age group 20-40 ")
print("2) Almost all people paid fare between 0-50")
```

```
Conculsion for above plots are :
1) Almost all people belongs to age group 20-40
2) Almost all people paid fare between 0-50
```

In [30]:
```
#1 represents survived while 0 represents dead
plt.subplot(2,2,1)
```

```
sns.boxplot(data=df,x='Sex',y='Age')
plt.subplot(2,2,2)
sns.boxplot(data=df,x='Survived',y='Age')
plt.subplot(2,2,3)
sns.boxplot(data=df,x='Sex',y='Fare')
plt.subplot(2,2,4)
sns.boxplot(data=df,x='Survived',y='Fare')
plt.tight_layout()
plt.show()
```

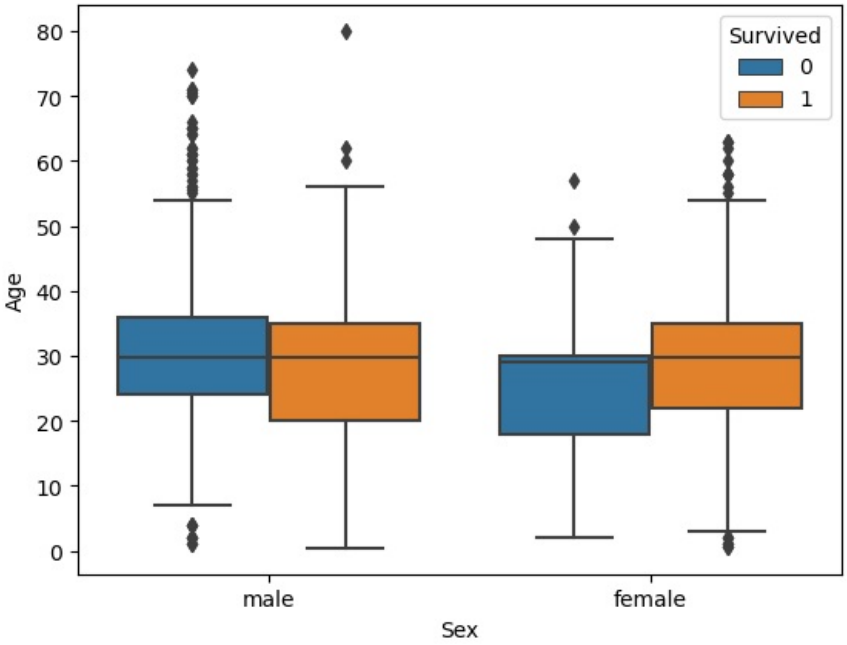


In [35]:
```
print("Conclusion from above plots are : ")
print("1) 75% males and females are in age group 30-40")
print("2) 75% people who died and survived belongs to age group 30-40")
print("3) More number of females had paid more fare compare to males ")
print("4) People who paid more fare has more survival rate ")
```

```
Conclusion from above plots are :
1) 75% males and females are in age group 30-40
2) 75% people who died and survived belongs to age group 30-40
3) More number of females had paid more fare compare to males
4) People who paid more fare has more survival rate
```

In [29]:
```
#Sex vs Age vs Survived
sns.boxplot(data=df,x='Sex',y='Age',hue='Survived')
```

Out[29]: <Axes: xlabel='Sex', ylabel='Age'>



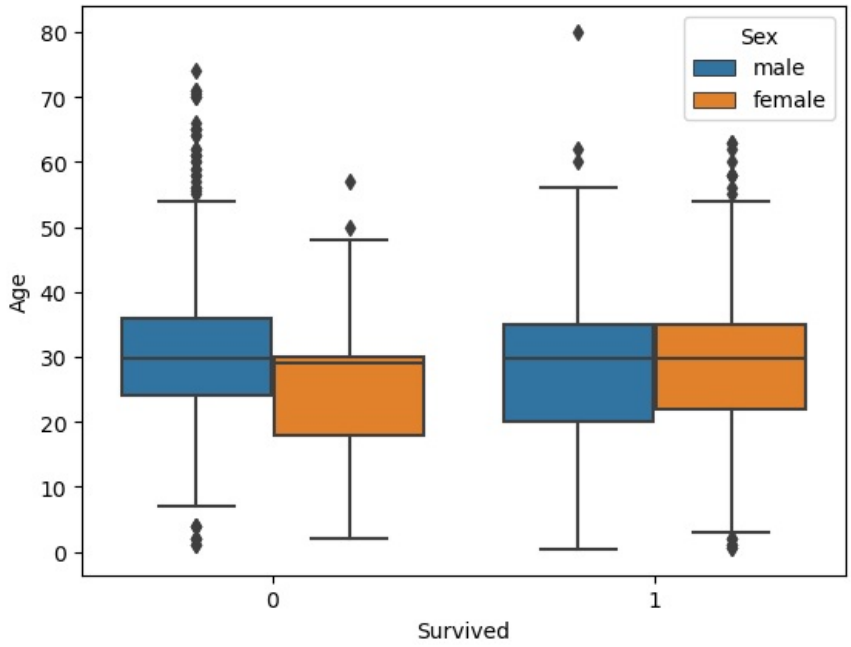
In [31]:
```
print("Conclusion from above is : ")
print("Less males survived compared to dead males in age group 30-40 ")
```

```
print("More females survived from age group 30-40")
```

```
Conclusion from above is :
Less males survived compared to dead males
More females survived comopared to dead females
```

In [32]:
```
sns.boxplot(data=df,x='Survived',y='Age',hue='Sex')
```

Out[32]: <Axes: xlabel='Survived', ylabel='Age'>



In [33]:
```
print('Conclusion from above plot is : ')
print("More number of Males died compared to females ")
```

```
Conclusion from above plot is :
More number of Males died compared to females
```