In [25]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [26]:
```
df=pd.read_csv('tested.csv')
```

In [27]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [28]:
```
df.describe()
```

Out[28]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	418.000000	418.000000	418.000000	332.000000	418.000000	418.000000	41
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	3
std	120.810458	0.481622	0.841838	14.181209	0.896760	0.981429	5
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	
25%	996.250000	0.000000	1.000000	21.000000	0.000000	0.000000	
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	0.000000	1
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	0.000000	3
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	51

In [29]:
```
df.head()
```

Out[29]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.(
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.(
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.(
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.

In [30]:
```
df.isnull()
```

Out[30]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
413	False	False	False	False	False	True	False	False	False	False
414	False	False	False	False	False	False	False	False	False	False
415	False	False	False	False	False	False	False	False	False	False
416	False	False	False	False	False	True	False	False	False	False
417	False	False	False	False	False	True	False	False	False	False

418 rows × 12 columns

In [31]:
```
df.isnull().sum()
```

Out[31]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

In [32]:
```
df['Age']=df['Age'].fillna(df['Age'].median())
```

In [33]:
```
df.isnull().sum()
```

Out[33]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

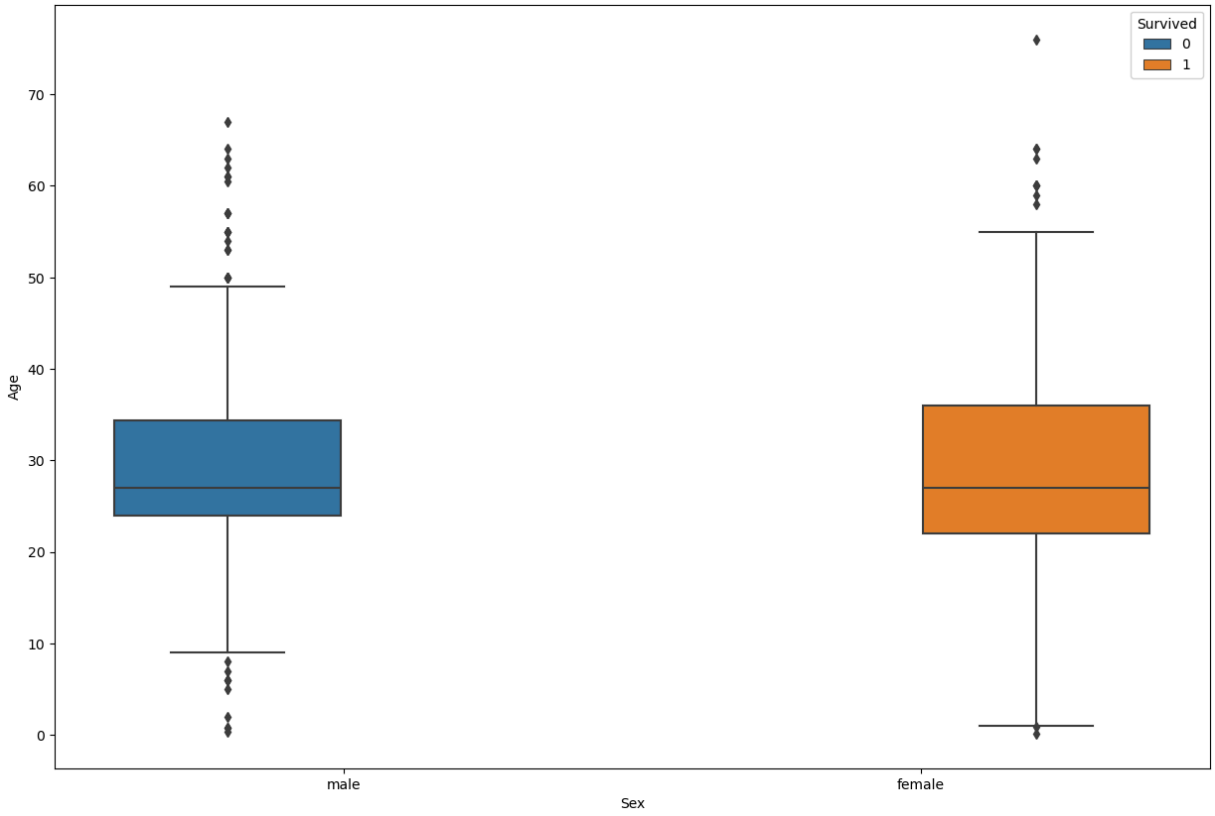
In [34]:
```
df['Fare']=df['Fare'].fillna(df['Fare'].mean())
```

In [35]:
```
df.isnull().sum()
```

Out[35]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          327
Embarked         0
dtype: int64
```

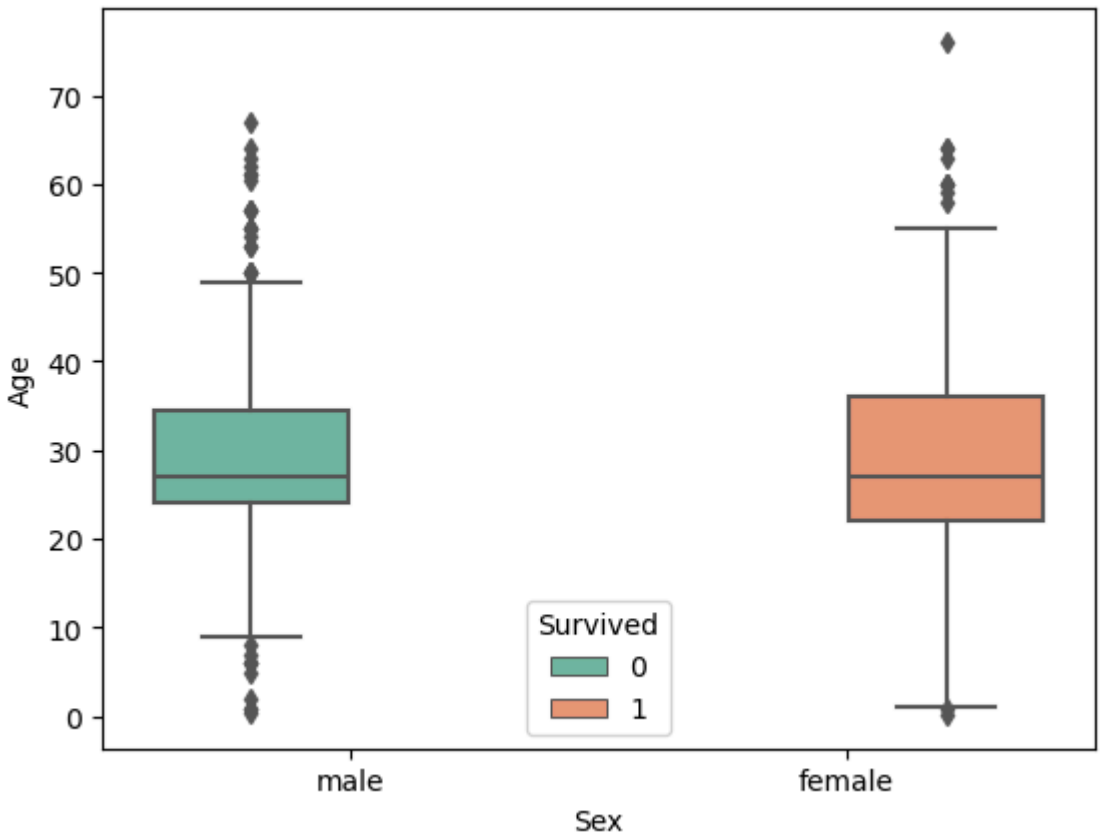
In [39]:
```
plt.figure(figsize=(15,10))
sns.boxplot(data=df, x="Sex", y="Age", hue="Survived")
```

Out[39]:
```
<Axes: xlabel='Sex', ylabel='Age'>
```



In [42]:
```
sns.boxplot(x = df['Sex'],
            y = df['Age'],
            hue = df['Survived'],
            palette = 'Set2')
```

Out[42]: <Axes: xlabel='Sex', ylabel='Age'>



In [ ]: