Untitled

February 14, 2025

[1]:
```
import pandas as pd
```

[2]:
```
import numpy as np
```

[4]:
```
import matplotlib.pyplot as plt
```

[5]:
```
import seaborn as sns
```

[8]:
```
df = pd.read_csv("Titanic.csv")
```

[9]:
```
df.head()
```

[9]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

[10]:
```
df.isnull().sum()
```

[10]:
```
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
```

[12]:
```
df.dtypes
```

[12]:
```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

[13]:
```
print(df.columns)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

[14]:
```
print(df.head())
print(df.info())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
```

```
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
<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
None
```

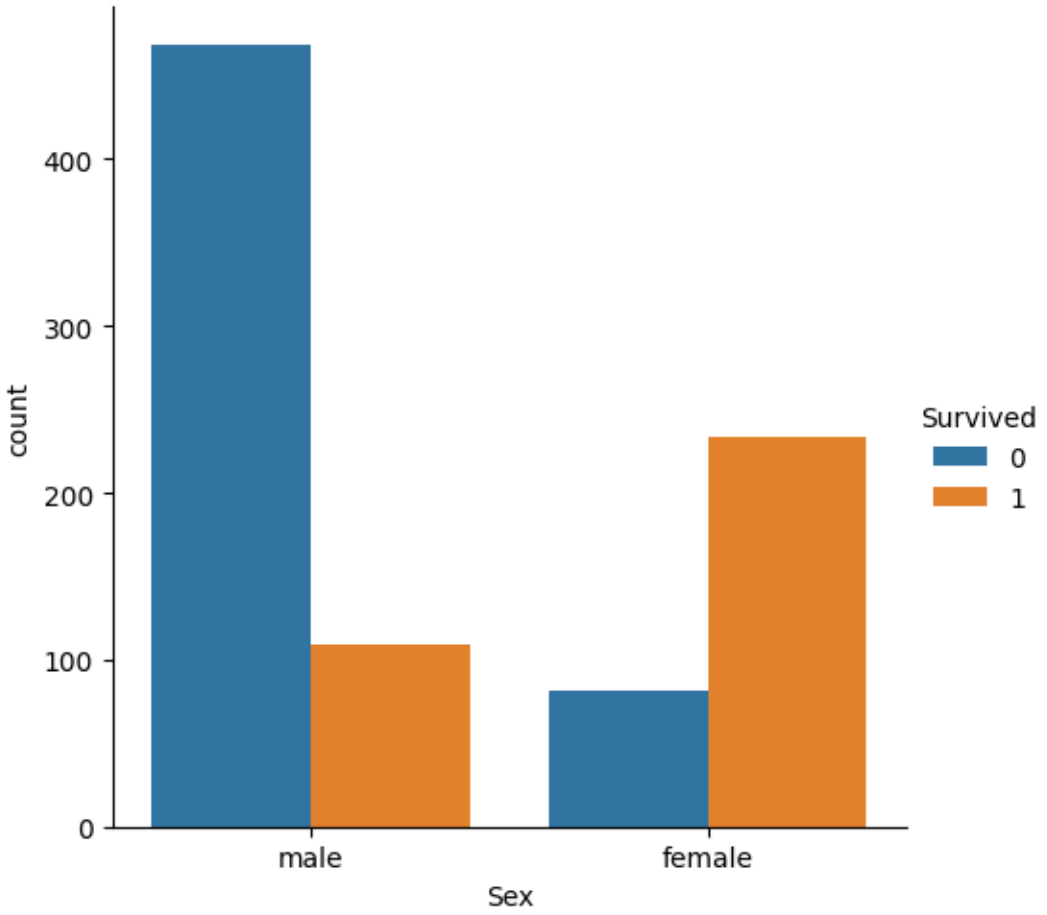
[15]:
```
type(df)
```

[15]:
```
pandas.core.frame.DataFrame
```

[16]:
```
df.columns = df.columns.str.strip()
```

[17]:
```
sns.catplot(x ='Sex', hue ='Survived' , kind = 'count',data =df)
```

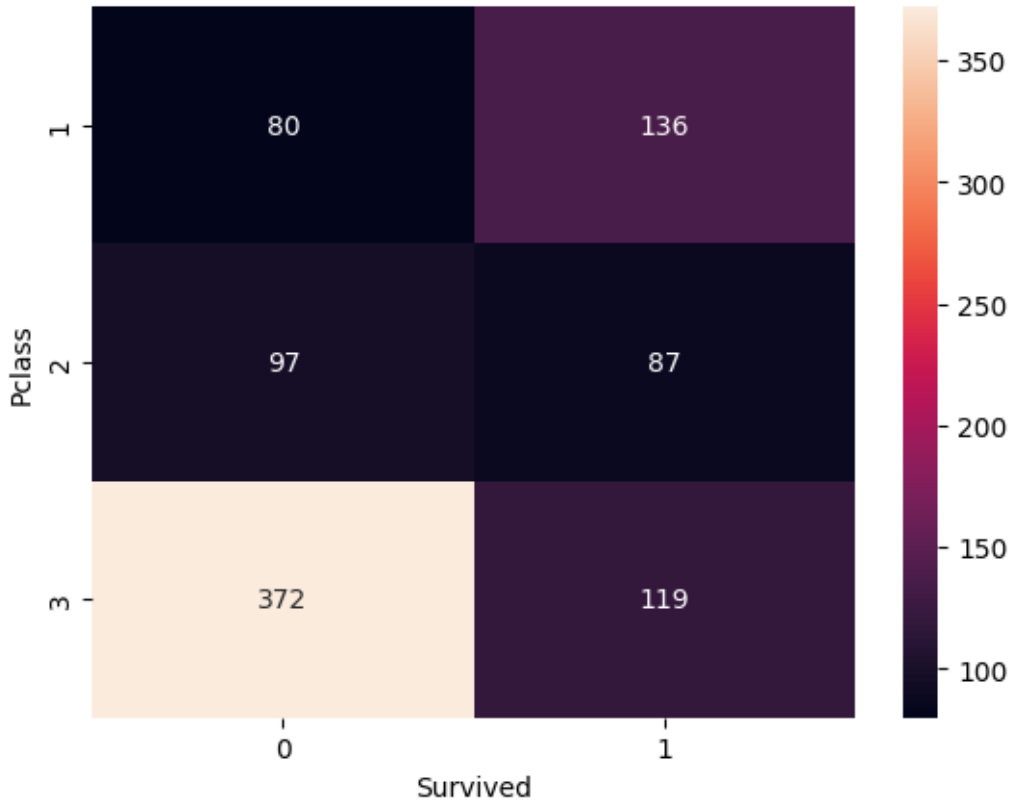
[17]:
```
<seaborn.axisgrid.FacetGrid at 0x7fd1051a9820>
```



[18]:
```
pclass_survived = df.groupby(['Pclass','Survived']).size().unstack()
```

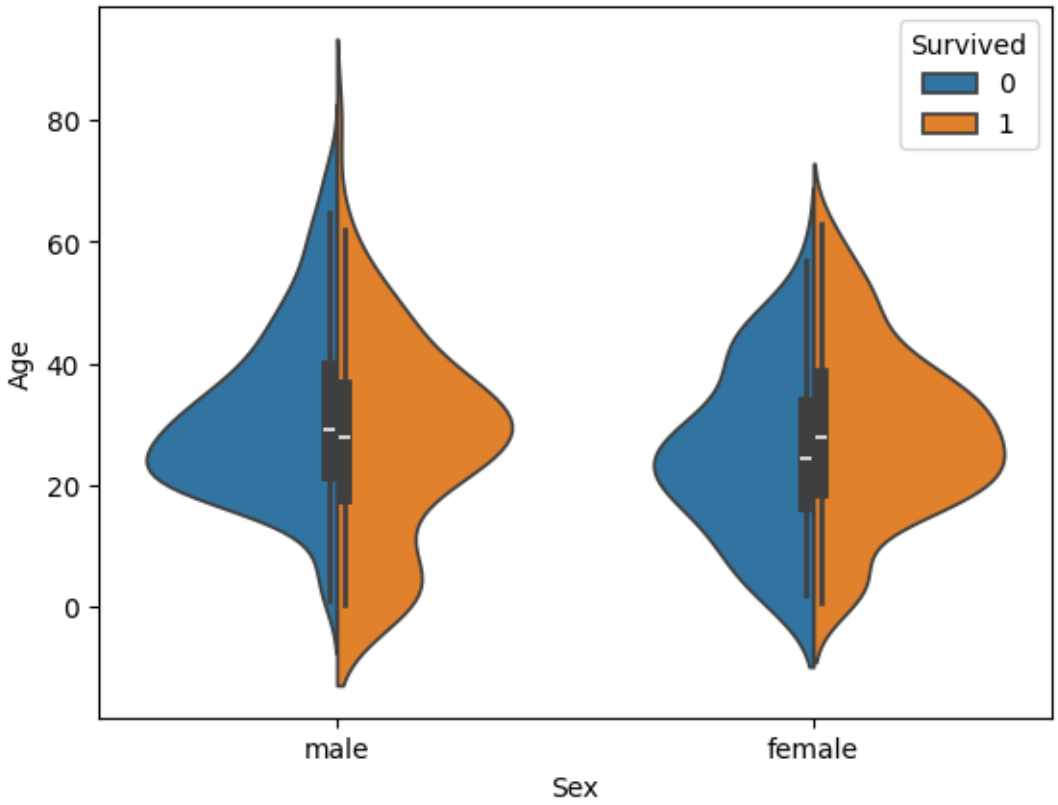
[19]:
```
sns.heatmap(pclass_survived,annot = True,fmt='d')
```

[19]: <Axes: xlabel='Survived', ylabel='Pclass'>



[20]:
```
sns.violinplot(x='Sex',y='Age',hue='Survived',data=df,split=True)
```

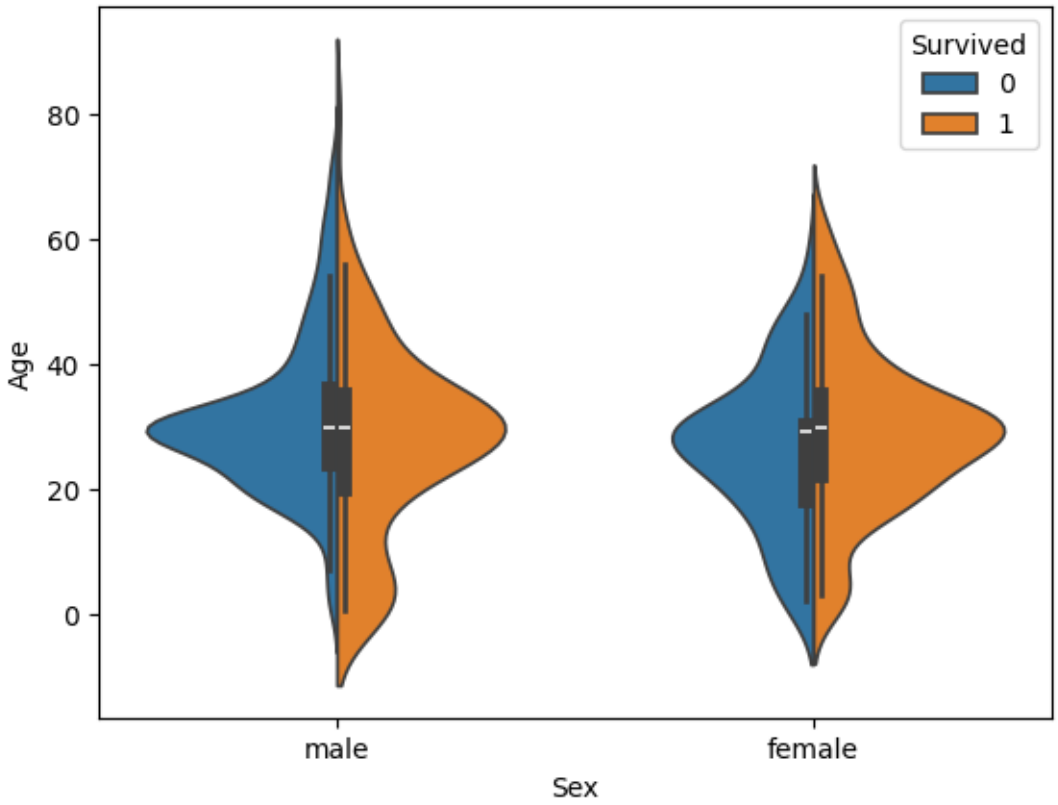
[20]: <Axes: xlabel='Sex', ylabel='Age'>



[21]:
```
df['Age'].fillna(df['Age'].mean(),inplace=True)
```

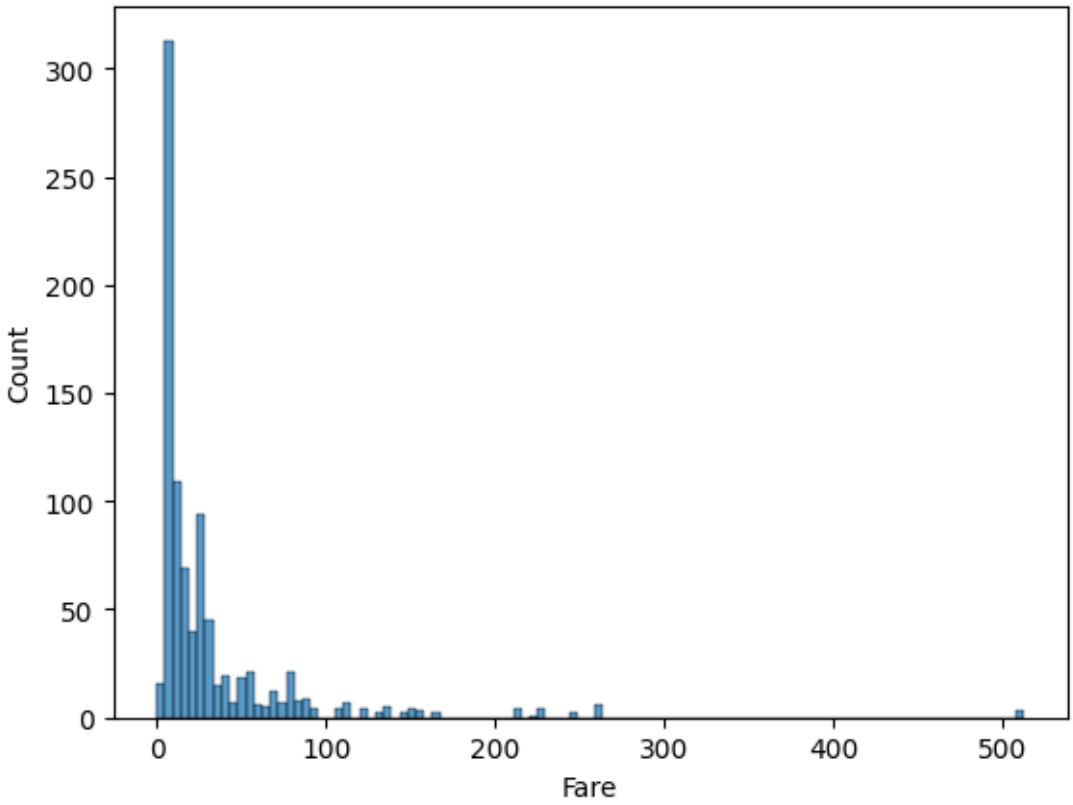
[22]:
```
sns.violinplot(x='Sex',y='Age',hue='Survived',data=df,split=True)
```

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



[23]:
```
sns.histplot(df['Fare'])
```

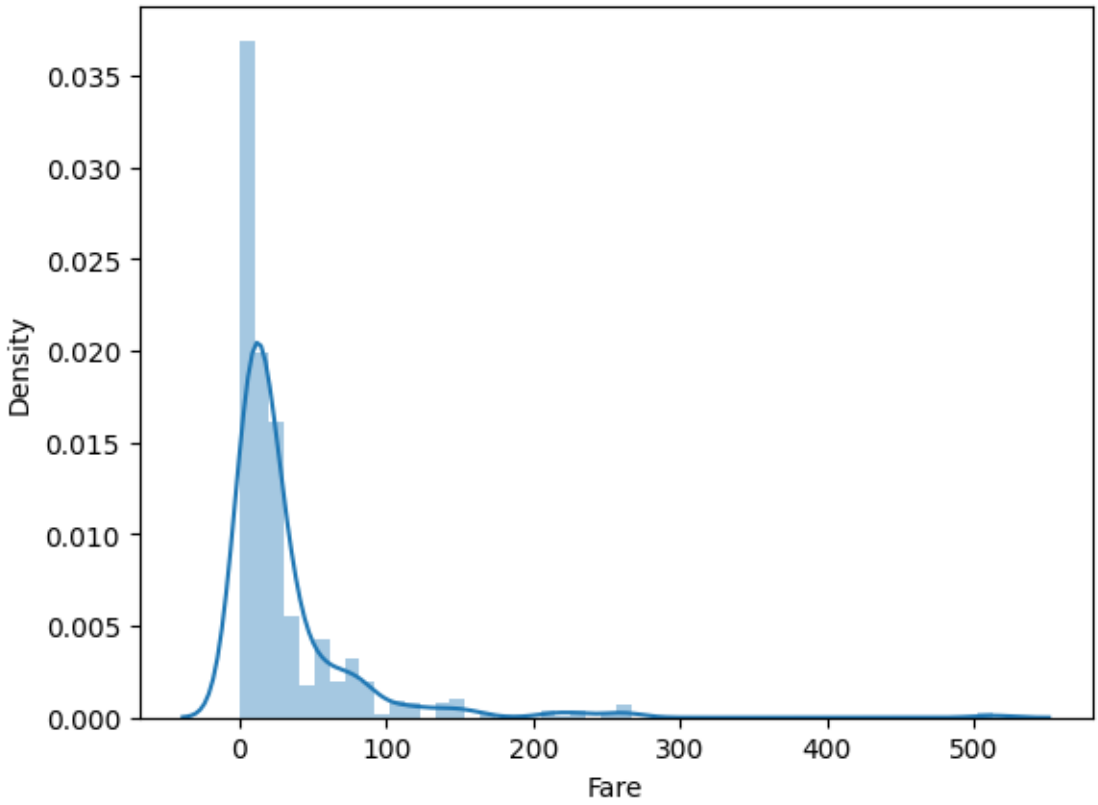
[23]:
```
<Axes: xlabel='Fare', ylabel='Count'>
```



[24]:
```
sns.distplot(df['Fare'])
```

```
/tmp/ipykernel_5225/3425841524.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['Fare'])
```

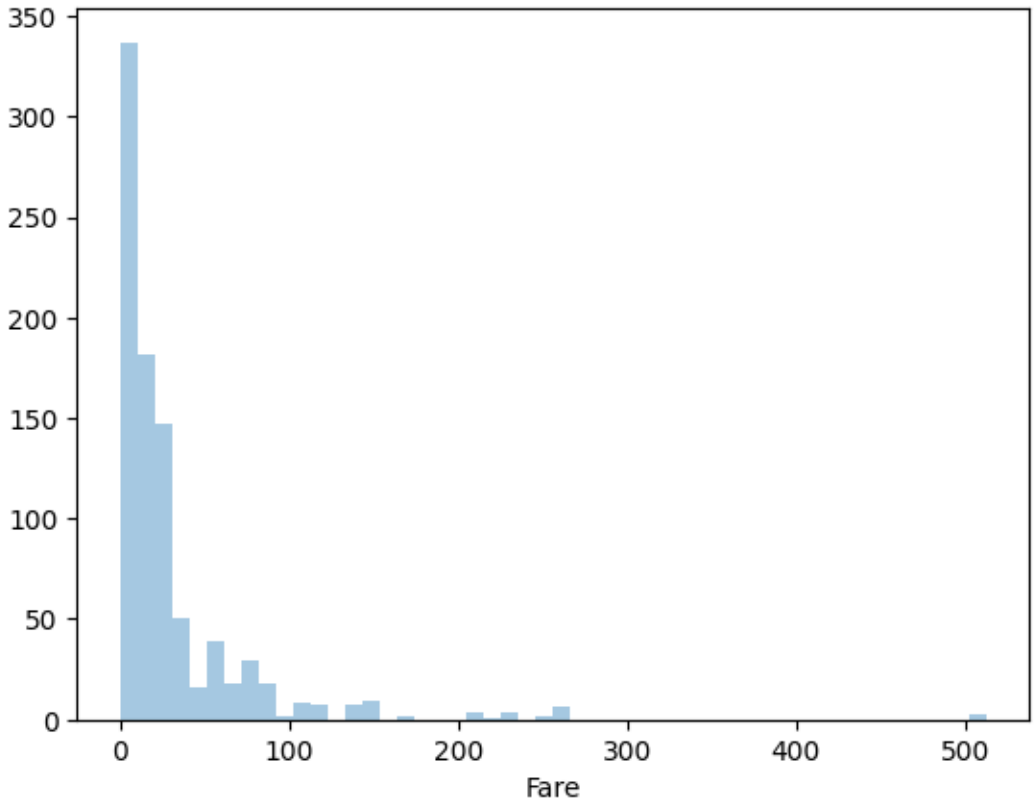
[24]: `<Axes: xlabel='Fare', ylabel='Density'>`



[25]:
```
sns.distplot(df['Fare'],kde=False)
```

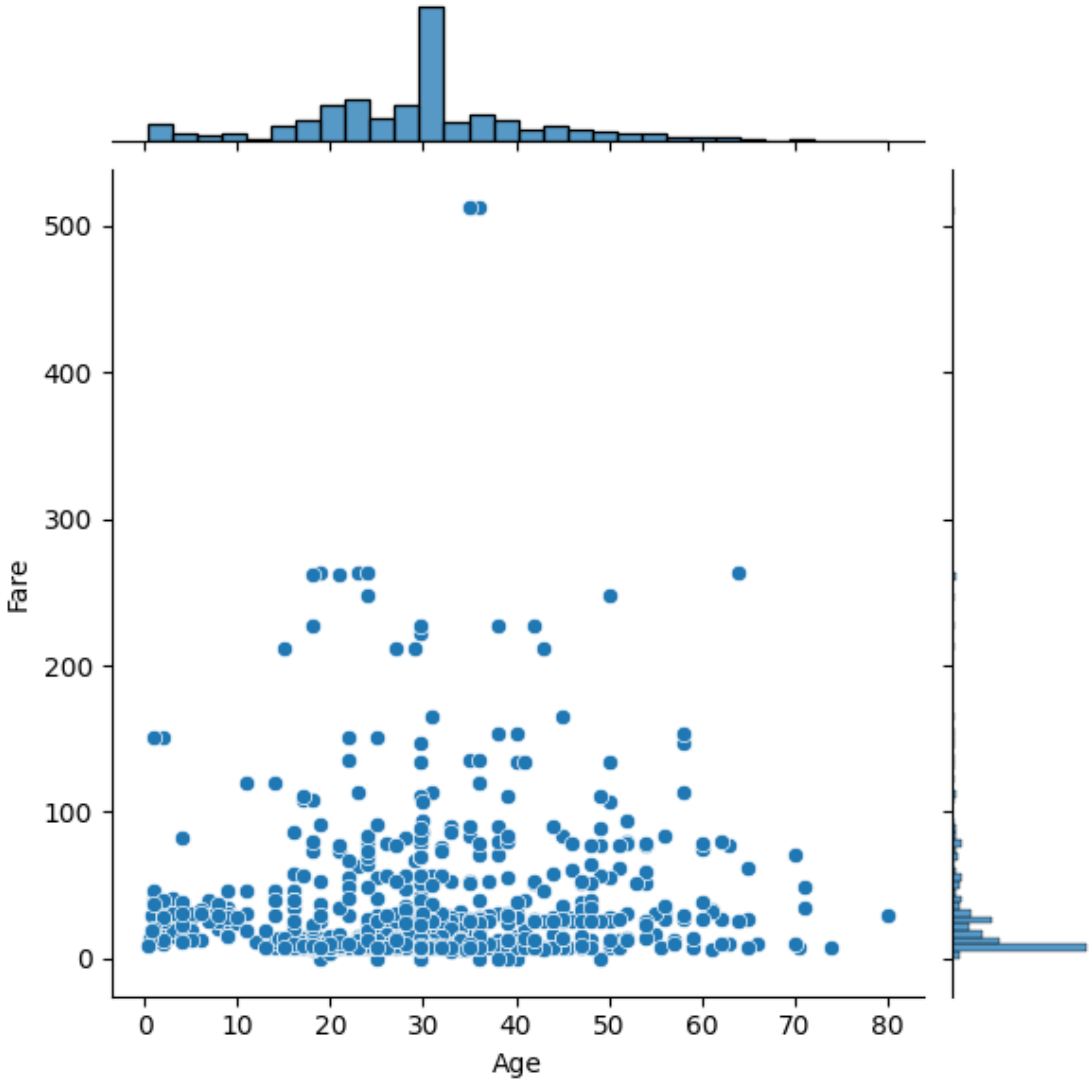
```
/tmp/ipykernel_5225/2475353197.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['Fare'],kde=False)
```

[25]: `<Axes: xlabel='Fare'>`



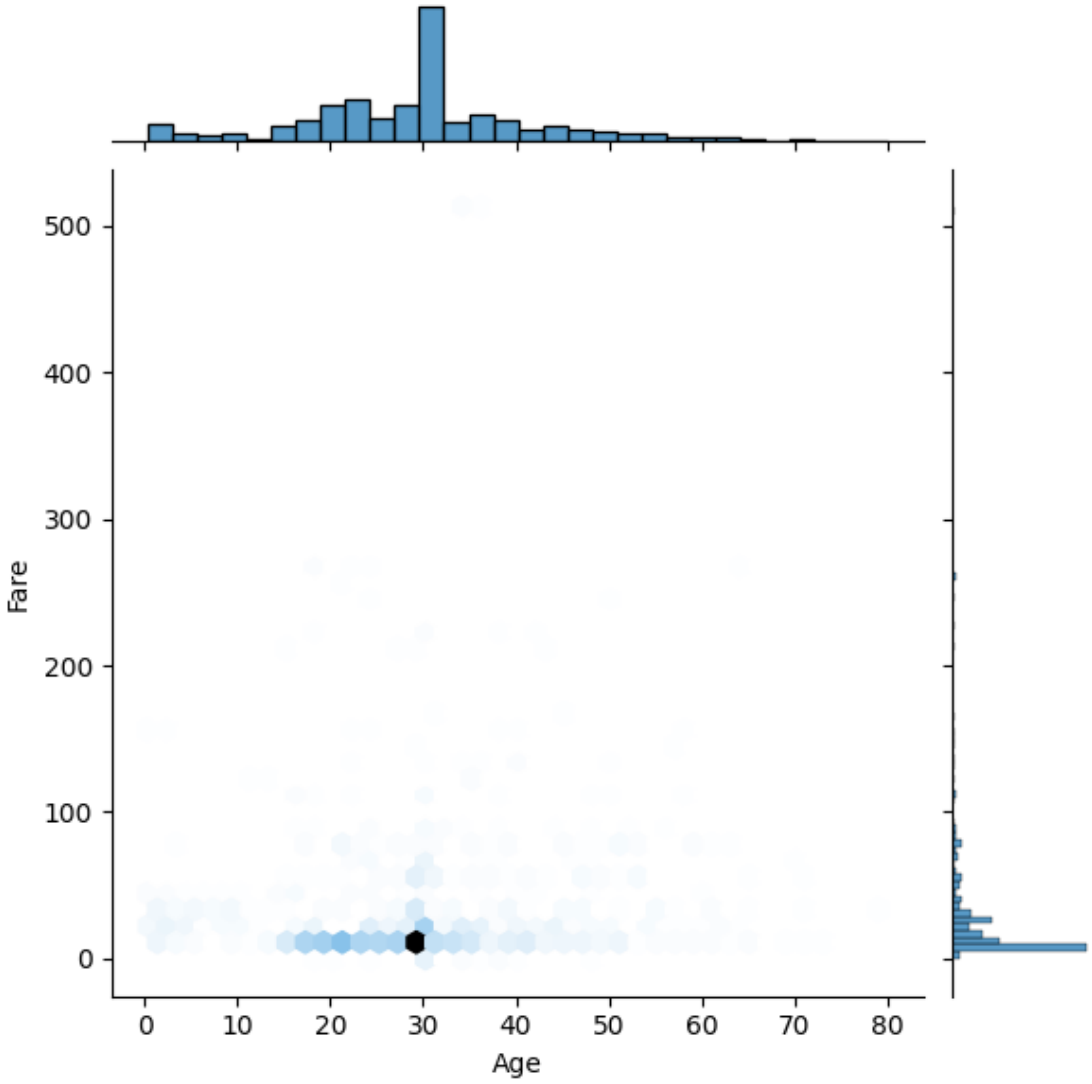
[27]:
```
sns.jointplot(x = df['Age'] ,y = df['Fare'] ,kind ='scatter')
```

[27]: <seaborn.axisgrid.JointGrid at 0x7fd100839f10>



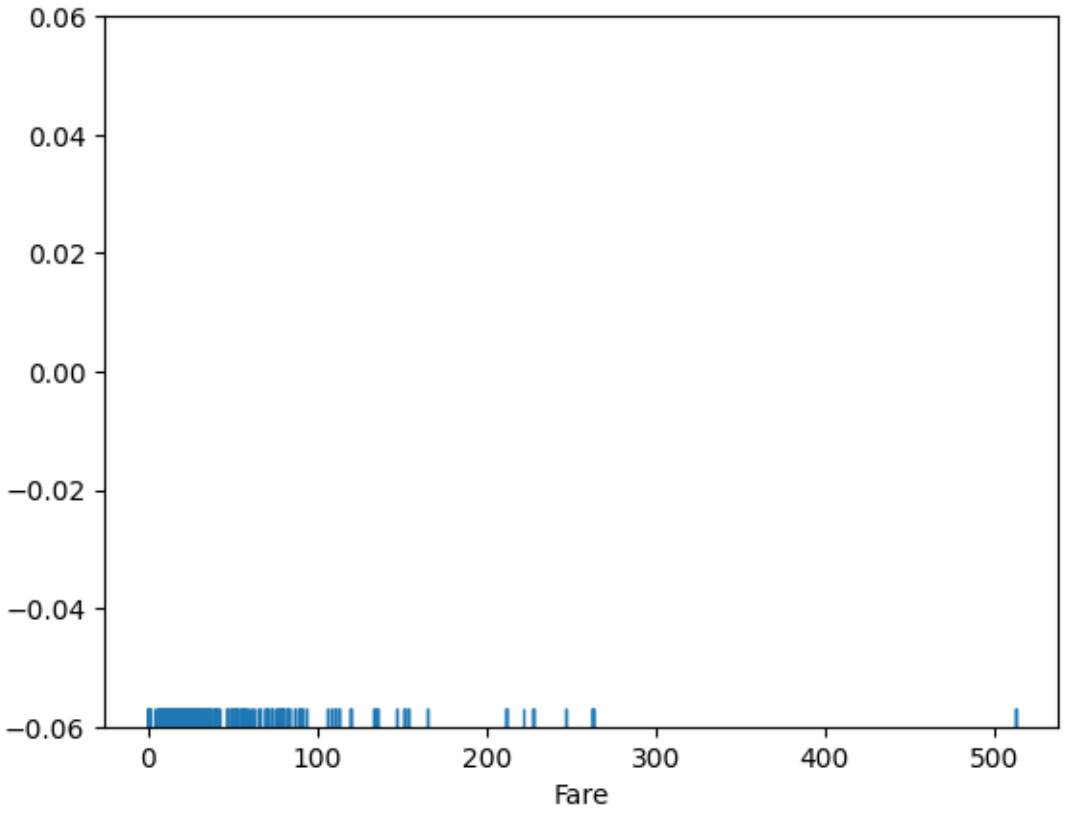
[28]:
```
sns.jointplot(x = df['Age'] ,y = df['Fare'] ,kind ='hex')
```

[28]: <seaborn.axisgrid.JointGrid at 0x7fd10059ba90>



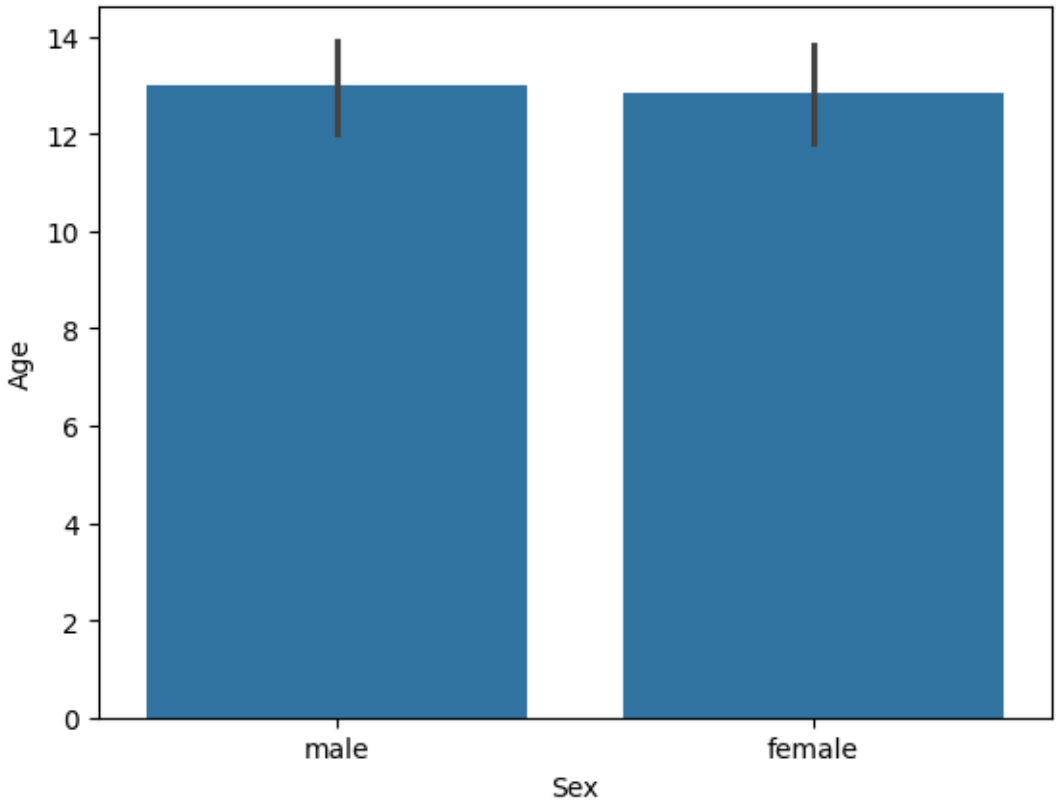
[29]:
```
sns.rugplot (df['Fare'])
```

[29]: <Axes: xlabel='Fare'>



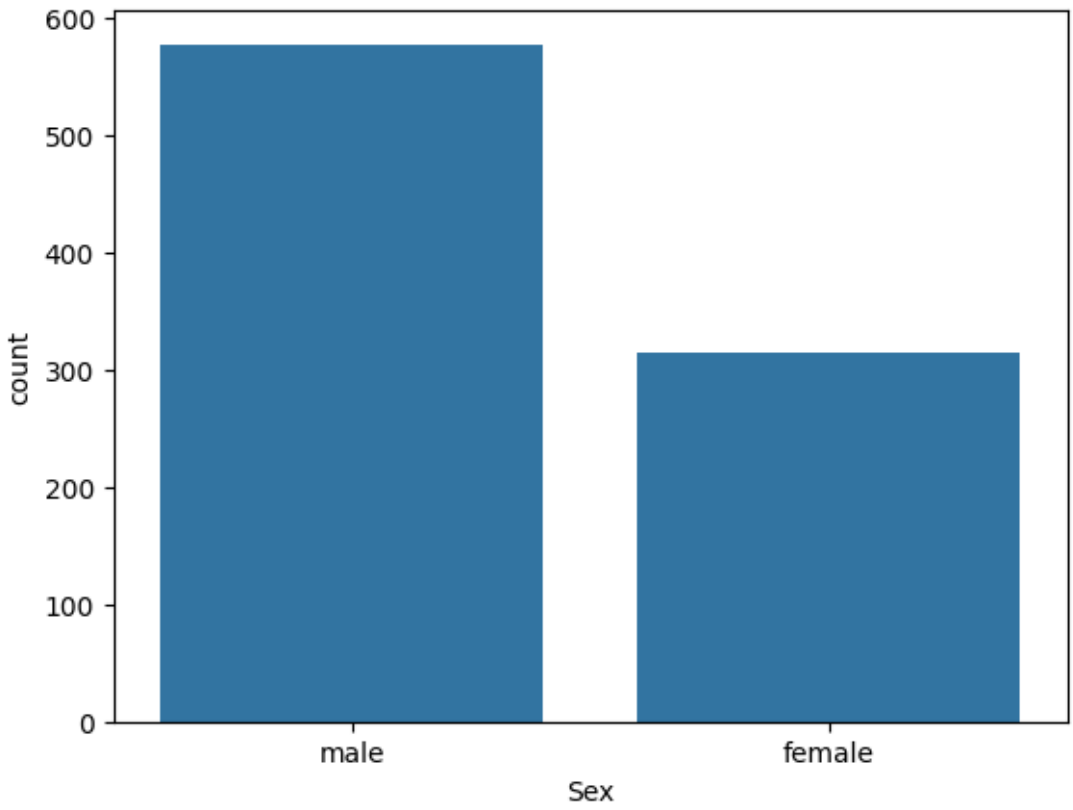
[30]:
```
sns.barplot(x = 'Sex' , y = 'Age',data = df ,estimator =np.std)
```

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



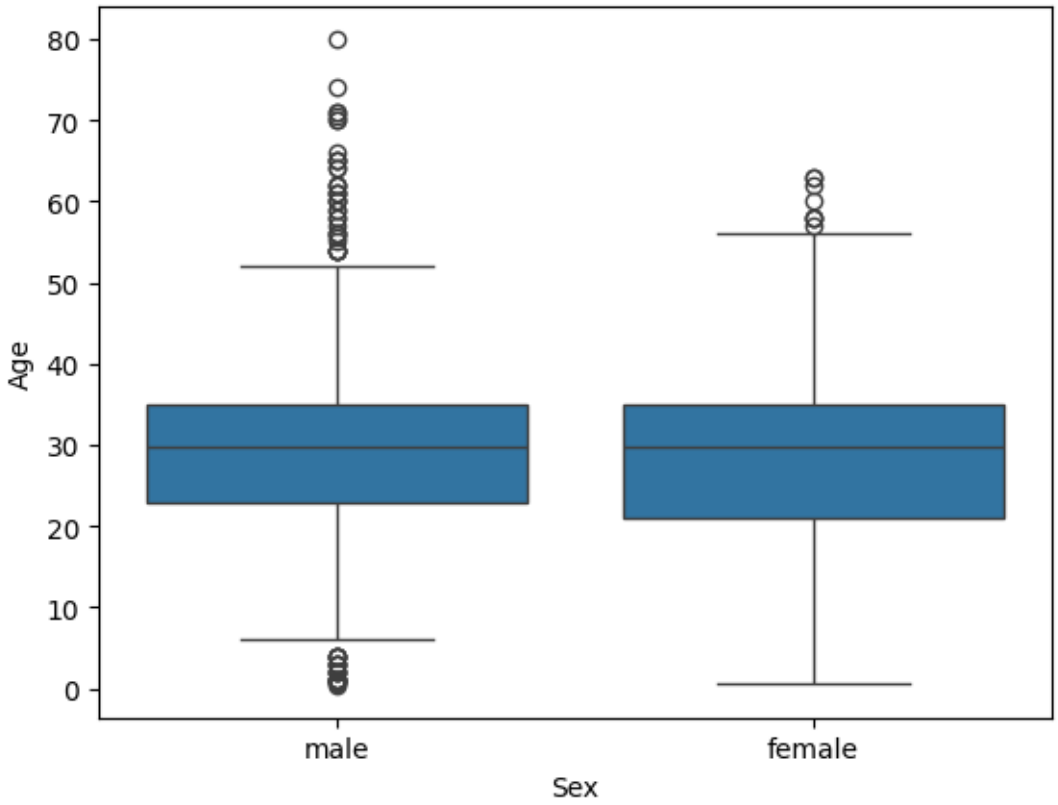
[31]:
```
sns.countplot(x = 'Sex' ,data = df)
```

[31]:
```
<Axes: xlabel='Sex', ylabel='count'>
```



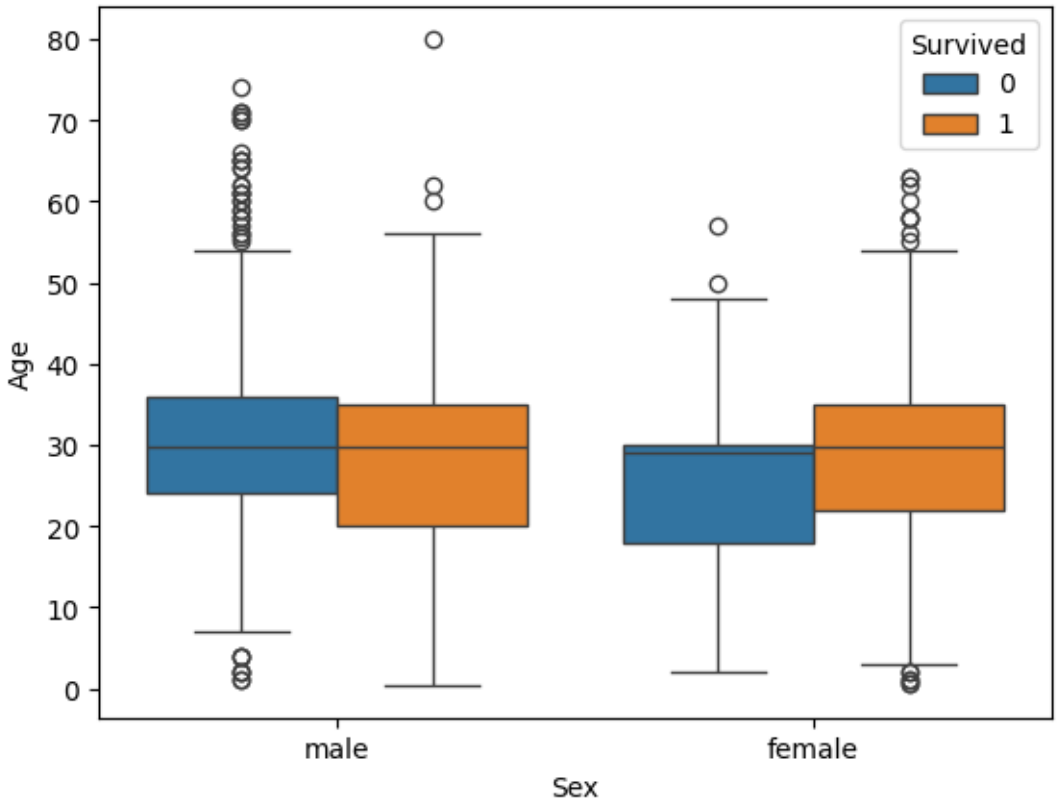
[33]:
```
sns.boxplot (x ='Sex',y = 'Age' ,data = df)
```

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



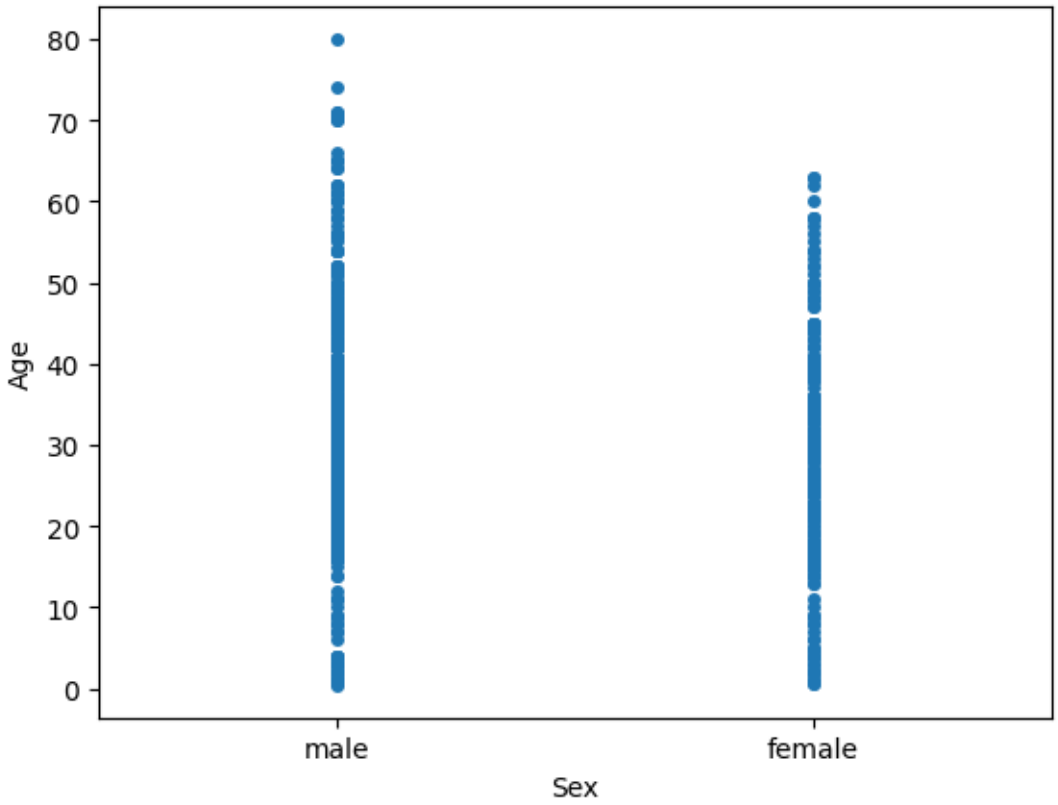
[34]:
```
sns.boxplot (x ='Sex',y = 'Age' ,data = df,hue="Survived")
```

[34]:
```
<Axes: xlabel='Sex', ylabel='Age'>
```



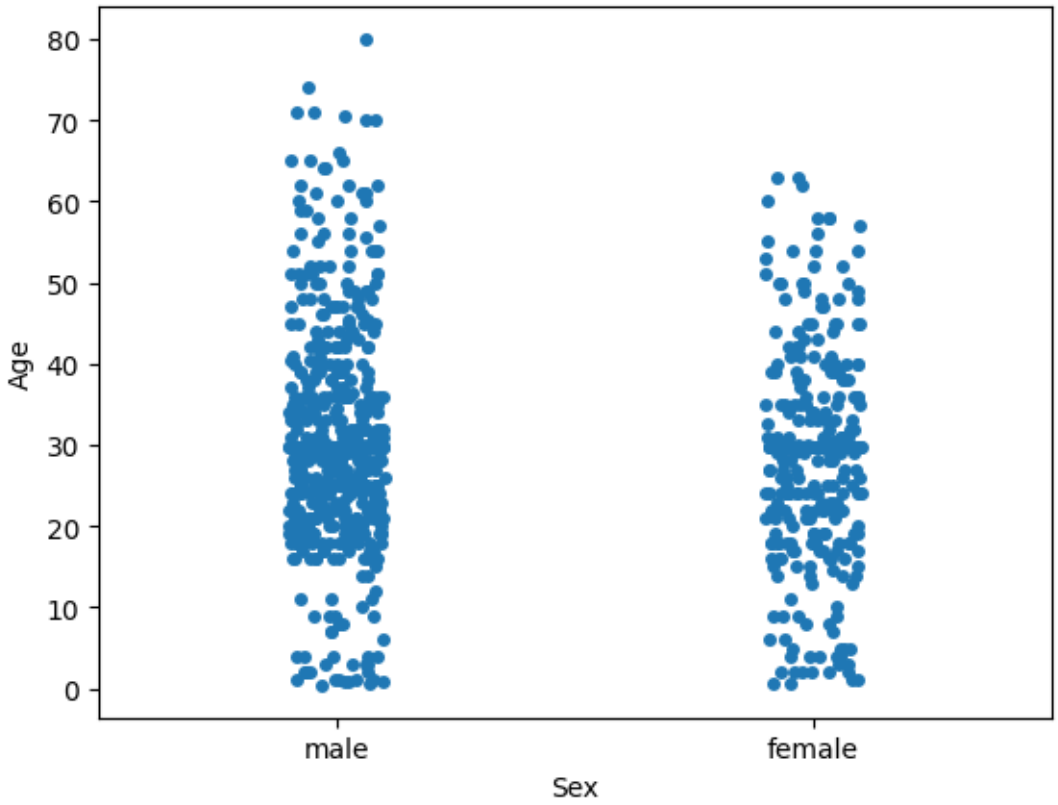
[35]:
```
sns.stripplot (x ='Sex',y = 'Age' ,data = df,jitter =False)
```

[35]:
```
<Axes: xlabel='Sex', ylabel='Age'>
```



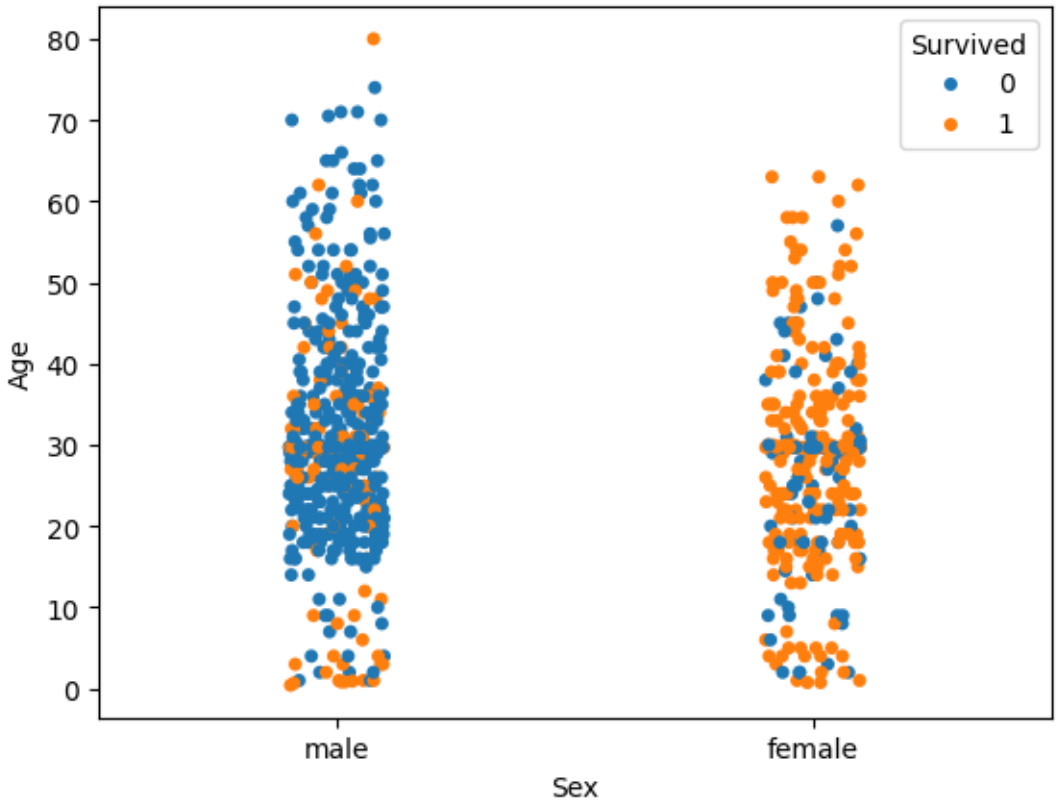
[36]:
```
sns.stripplot (x ='Sex',y = 'Age' ,data = df,jitter =True)
```

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



[37]:
```
sns.stripplot (x ='Sex',y = 'Age' ,data = df,jitter =True ,hue ='Survived')
```

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


Survived
0
1
Age
80
70
60
50
40
30
20
10
0
male
female
Sex

[ ]: