Assignment-8

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
data = pd.read_csv('https://raw.githubusercontent.com/dphi-official/Datasets/
 ↪master/titanic_data.csv')
data
```

[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	1
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4	5	0	3				
..	…	…	…				
886	887	0	2				
887	888	1	1				
888	889	0	3				
889	890	1	1				
890	891	0	3				

```
4                              Allen, Mr. William Henry
                                    male 35.0  0

     ..                                                 …
     …                                                  …   …

886                              Montvila, Rev. Juozas
                                    male 27.0  0

887                              Graham, Miss. Margaret
                                 Edith female 19.0  0

888                              Johnston, Miss.
                                 Catherine Helen
                                 "Carrie" female    NaN
                                    1

889                              Behr, Mr. Karl Howell
                                    male 26.0  0

890                              Dooley, Mr. Patrick
                                    male 32.0  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
     ..    …               …      …   …      …
     886     0           211536 13.0000  NaN       S
     887     0           112053 30.0000  B42       S
     888     2       W./C. 6607 23.4500  NaN       S
     889     0           111369 30.0000 C148       C
     890     0           370376  7.7500  NaN       Q
     [891 rows x 12 columns]
```

[3]:
```
data.shape
```

[3]:
```
(891, 12)
```

[4]:
```
data.describe()
```

[4]:
```
       PassengerId   Survived     Pclass        Age      SibSp \
count 891.000000 891.000000 891.000000 714.000000 891.000000
mean   446.000000   0.383838   2.308642  29.699118   0.523008
std    257.353842   0.486592   0.836071  14.526497   1.102743
min      1.000000   0.000000   1.000000   0.420000   0.000000
25%    223.500000   0.000000   2.000000  20.125000   0.000000
50%    446.000000   0.000000   3.000000  28.000000   0.000000
```

```
75%    668.500000   1.000000    3.000000  38.000000   1.000000
max    891.000000   1.000000    3.000000  80.000000   8.000000

           Parch       Fare
count          891.000000
891.000000  mean  0.381594
32.204208   std   0.806057
49.693429   min   0.000000
0.000000   25%    0.000000
7.910400
50%     0.000000   14.454200
75%     0.000000   31.000000
max     6.000000 512.329200
```

[5]:
```
data.describe(include = 'object')
```

[5]:
```
                    Name Sex Ticket Cabin Embarked
count                891   891    891204    889
unique               891     2    681147      3
top    Mangan, Miss. Mary male      G6      S
       347082
freq                   1 577       7  4    644
```

[6]:
```
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
```

[7]:
```
data['Age'] = data['Age'].fillna(np.mean(data['Age']))
```

[8]:
```
data['Cabin'] = data['Cabin'].fillna(data['Cabin'].mode()[0])
```

[9]:
```
data['Embarked'] = data['Embarked'].fillna(data['Embarked'].mode()[0])
```

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

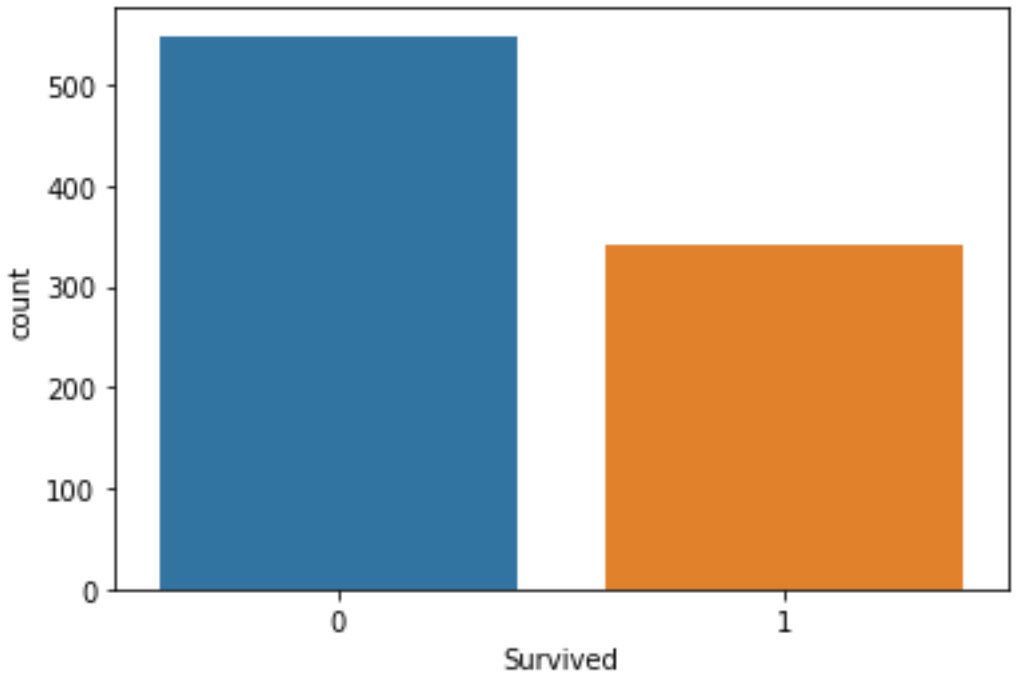
[10]:
```
PassengerId    0
Survived       0
```

[6]:
```
data.isnull().sum()
```

```
Pclass        0
Name          0
Sex           0
Age           0
SibSp         0
Parch         0
Ticket        0
Fare          0
Cabin         0
Embarked      0
dtype: int64
```

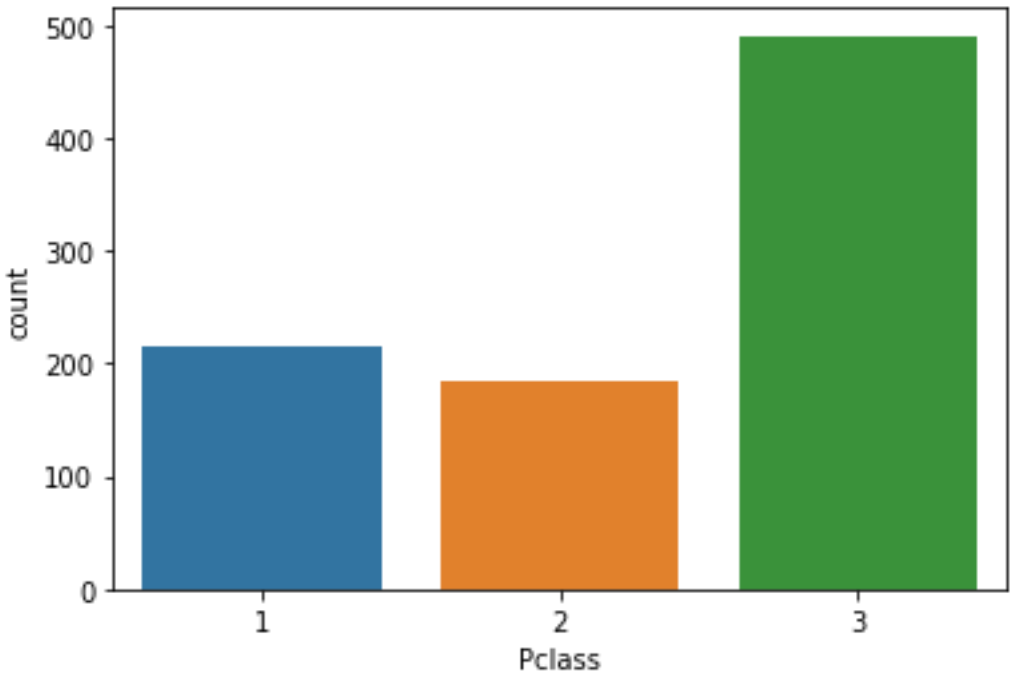
[11]:
```
sns.countplot(data['Survived'])
```

```
[11]: <matplotlib.axes._subplots.AxesSubplot at
0x1b5c823e7c0>
```



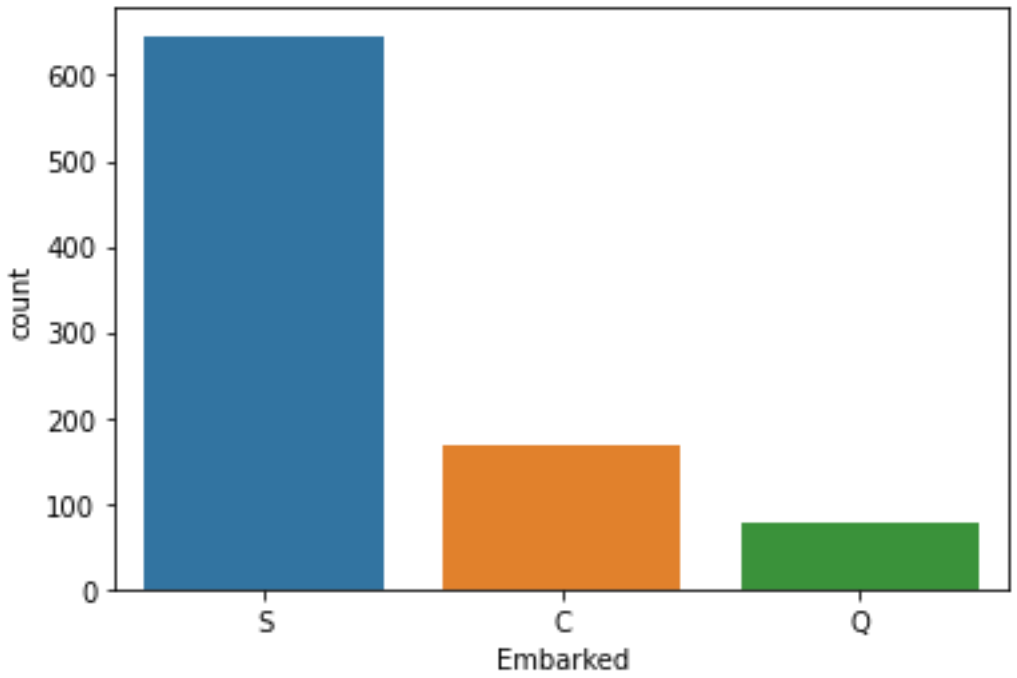
[12]:
```
sns.countplot(data['Pclass'])
```

```
[12]: <matplotlib.axes._subplots.AxesSubplot at
0x1b5c8997c70>
```


500
400
300
200
100
0
count
1
2
3
Pclass

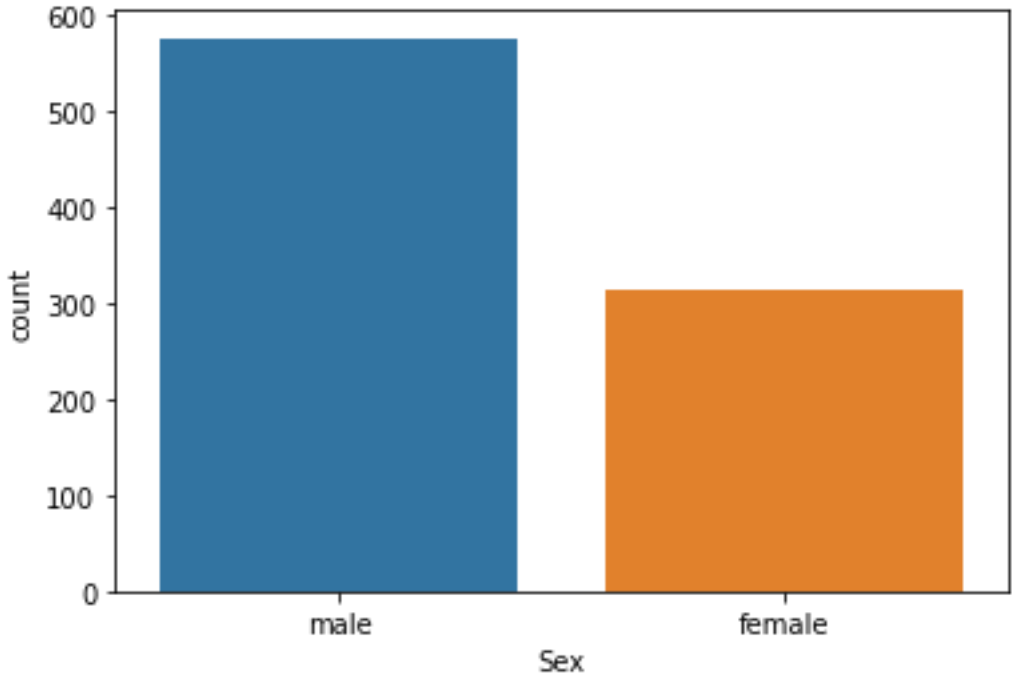
[13]: sns.countplot(data['Embarked'])

[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c89f9910>



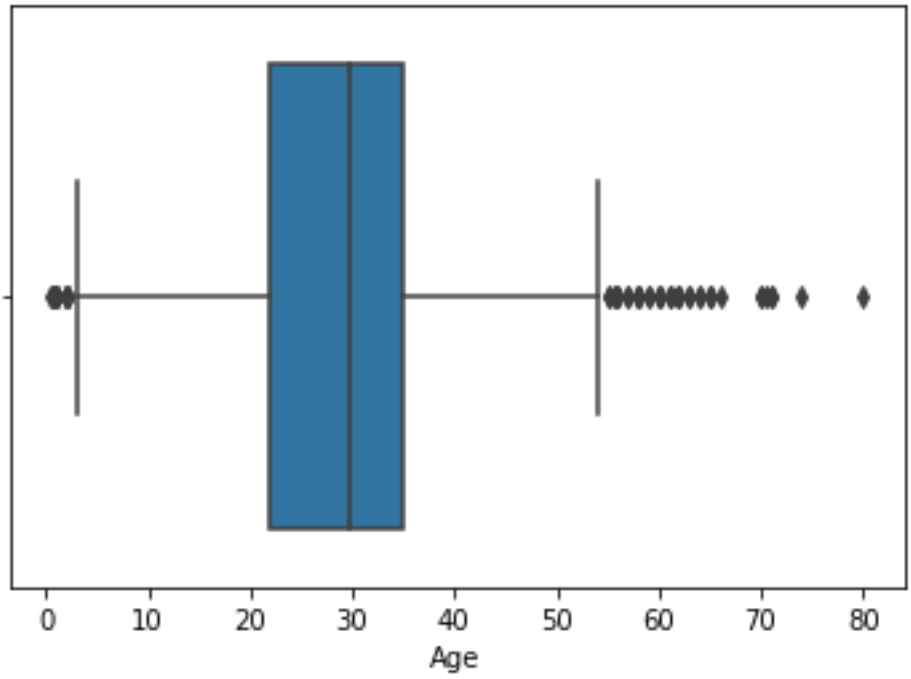
[14]: sns.countplot(data['Sex'])

[14]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8a4e8b0>



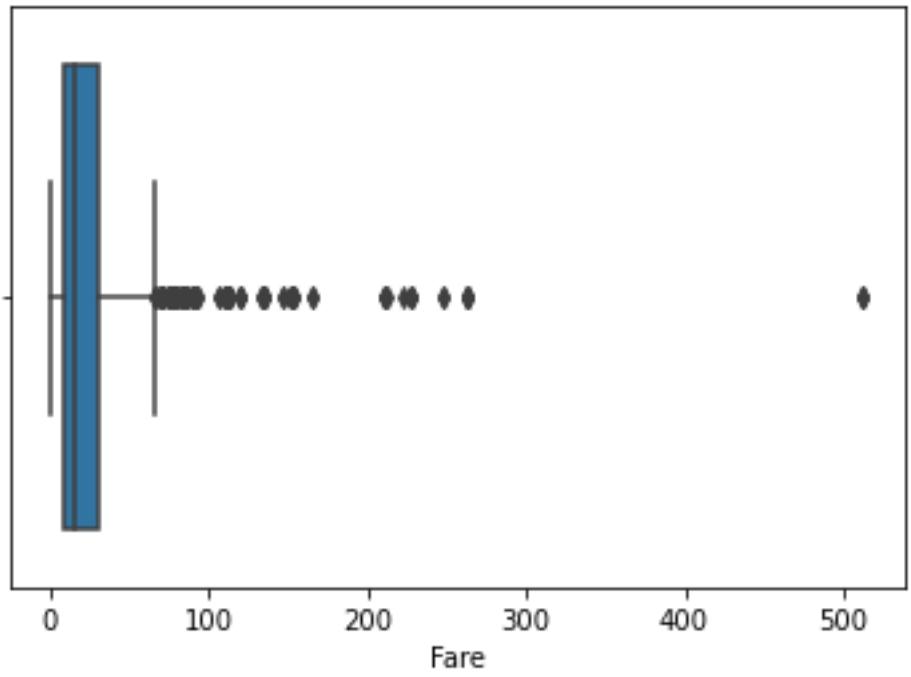
[15]:
```
sns.boxplot(data['Age'])
```

[15]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8a9c490>



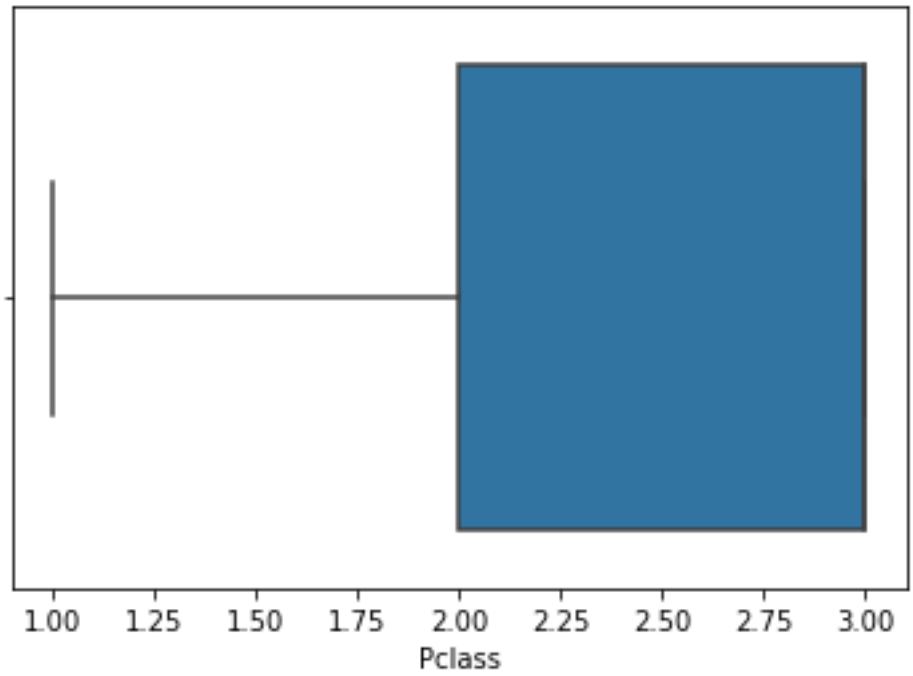
[16]: sns.boxplot(data['Fare'])

[16]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8aed7c0>



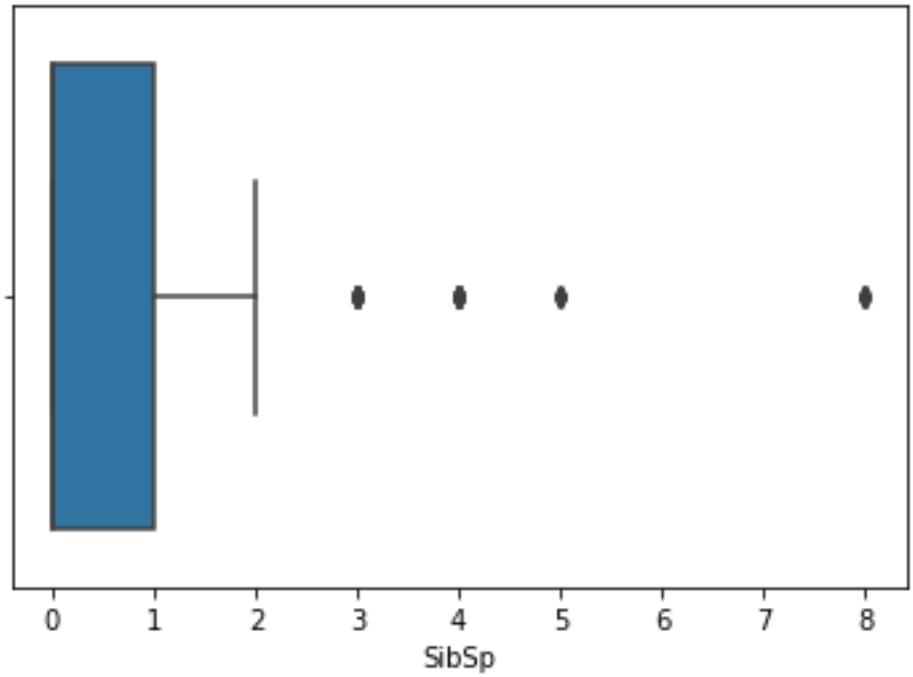
[17]:
```
sns.boxplot(data['Pclass'])
```

[17]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8b52c10>



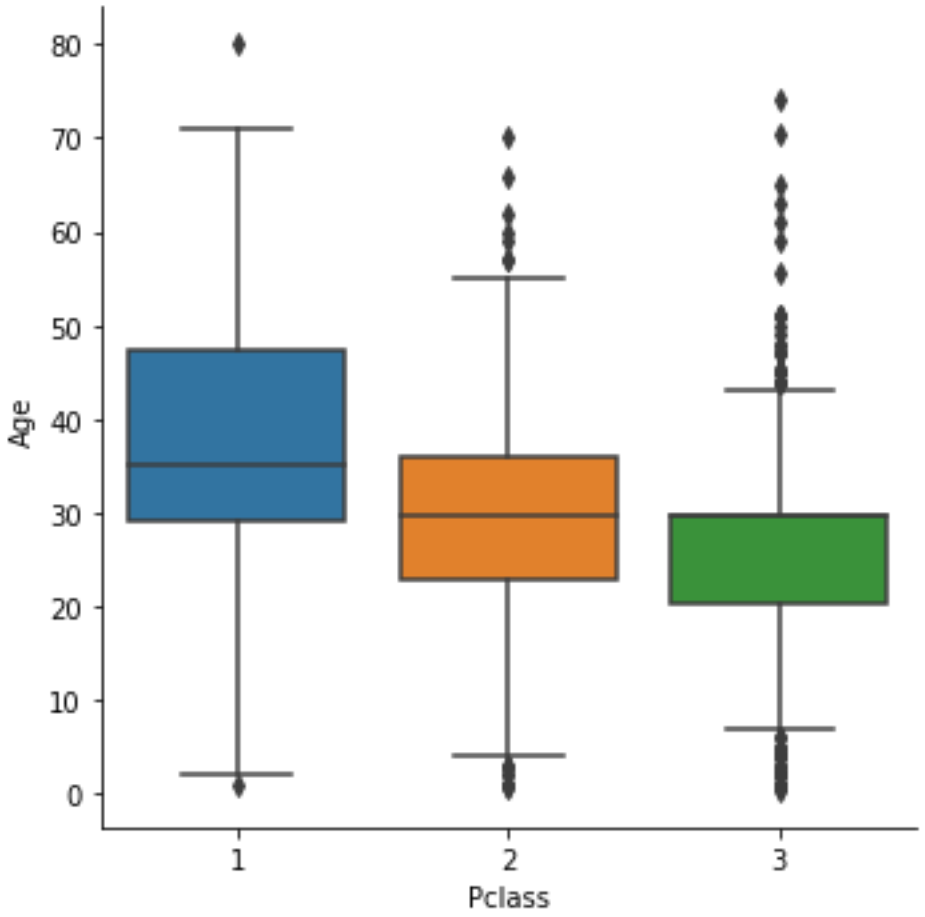
[18]:
```
sns.boxplot(data['SibSp'])
```

[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8b9f940>



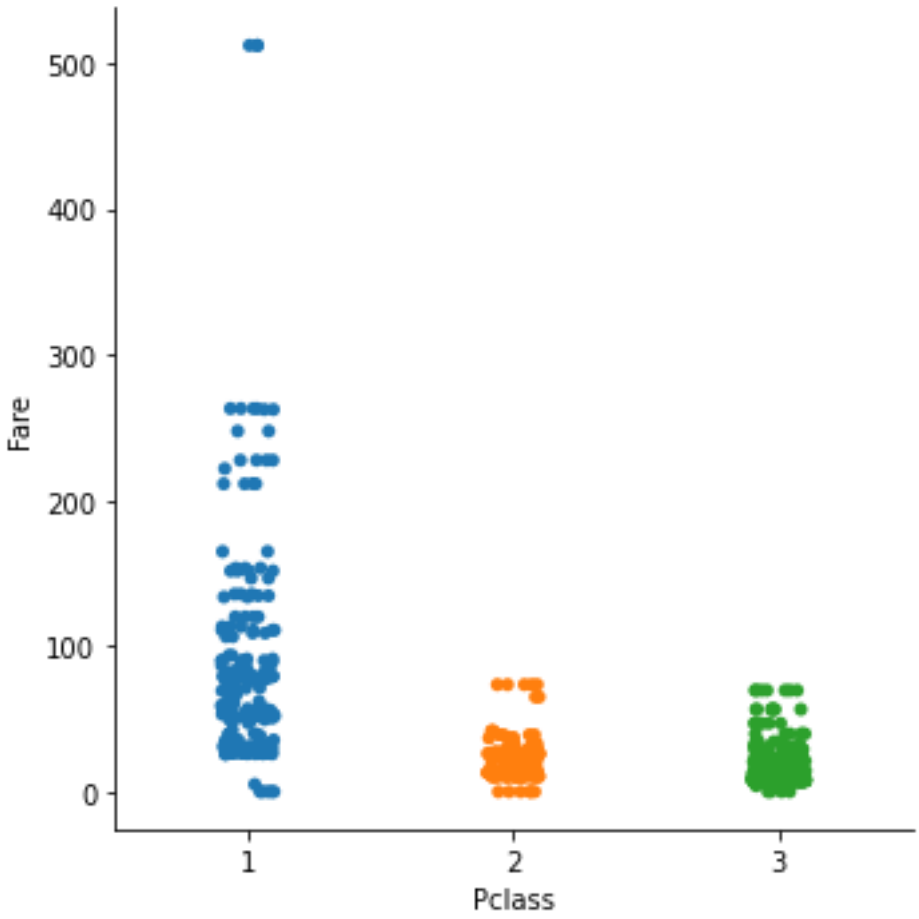
[19]:
```
sns.catplot(x= 'Pclass', y = 'Age', data=data, kind =
'box')
```

[19]:
```
<seaborn.axisgrid.FacetGrid at 0x1b5c8affca0>
```



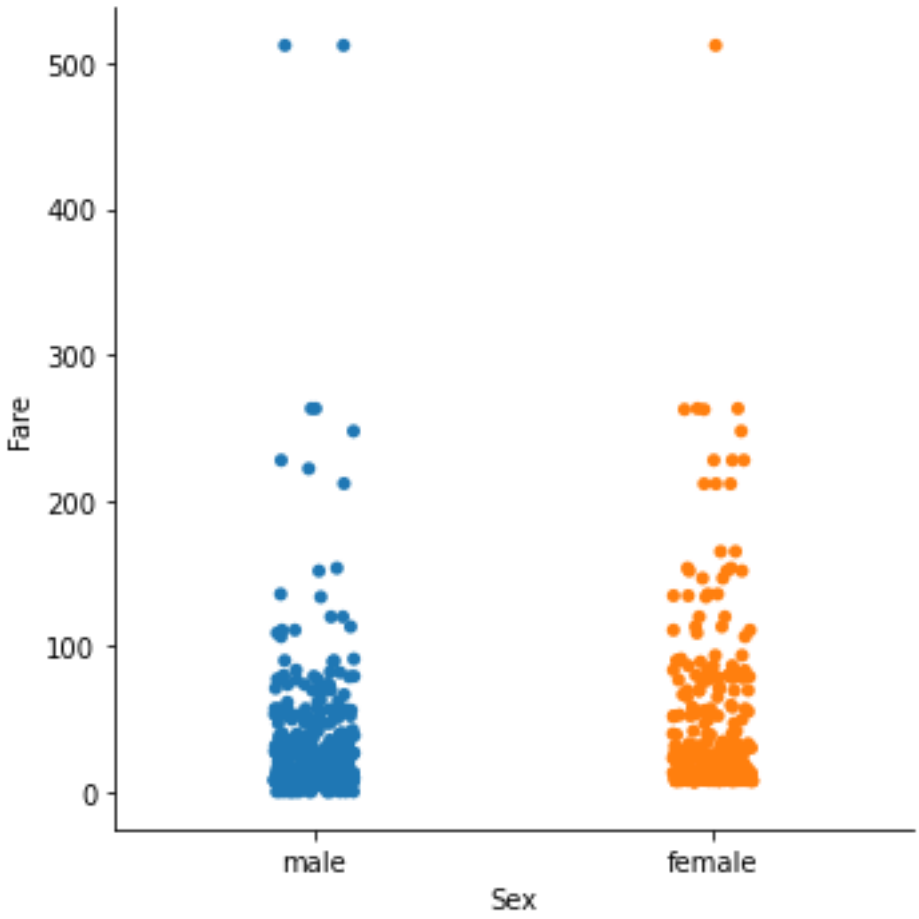
[20]:
```
sns.catplot(x= 'Pclass', y = 'Fare', data=data, kind =
'strip')
```

[20]: <seaborn.axisgrid.FacetGrid at 0x1b5c8ca3ee0>



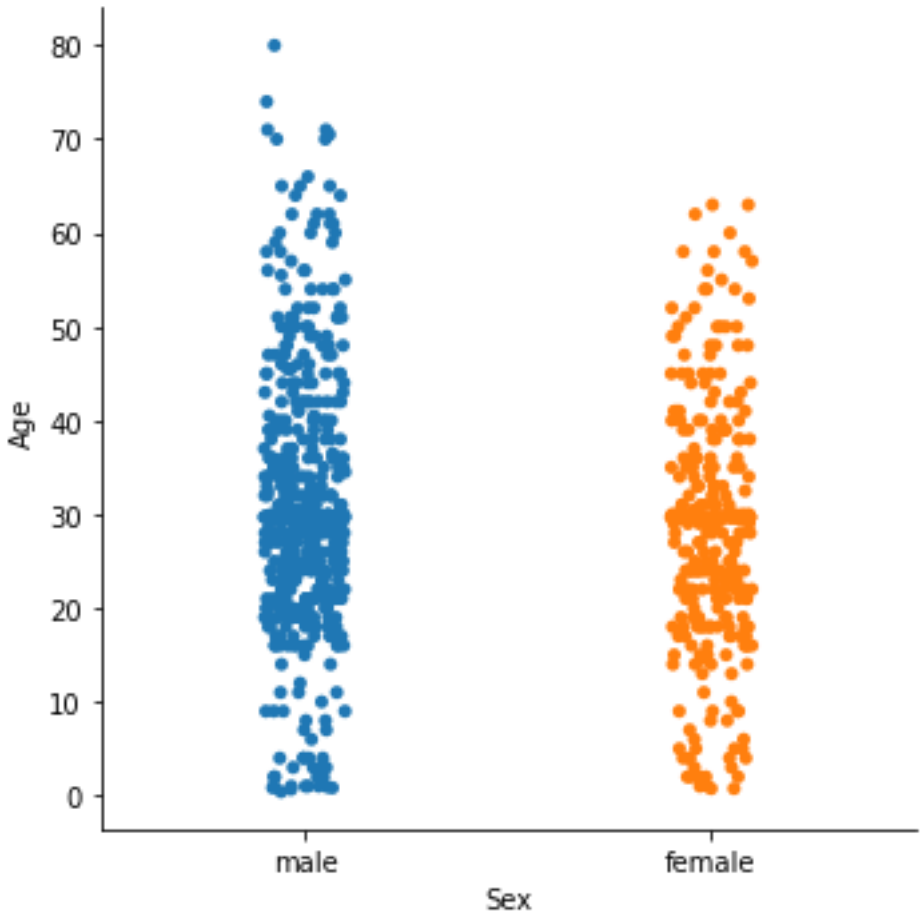

```
[21]: sns.catplot(x= 'Sex', y = 'Fare', data=data, kind =
'strip')
```

```
[21]: <seaborn.axisgrid.FacetGrid at 0x1b5c8aed9a0>
```



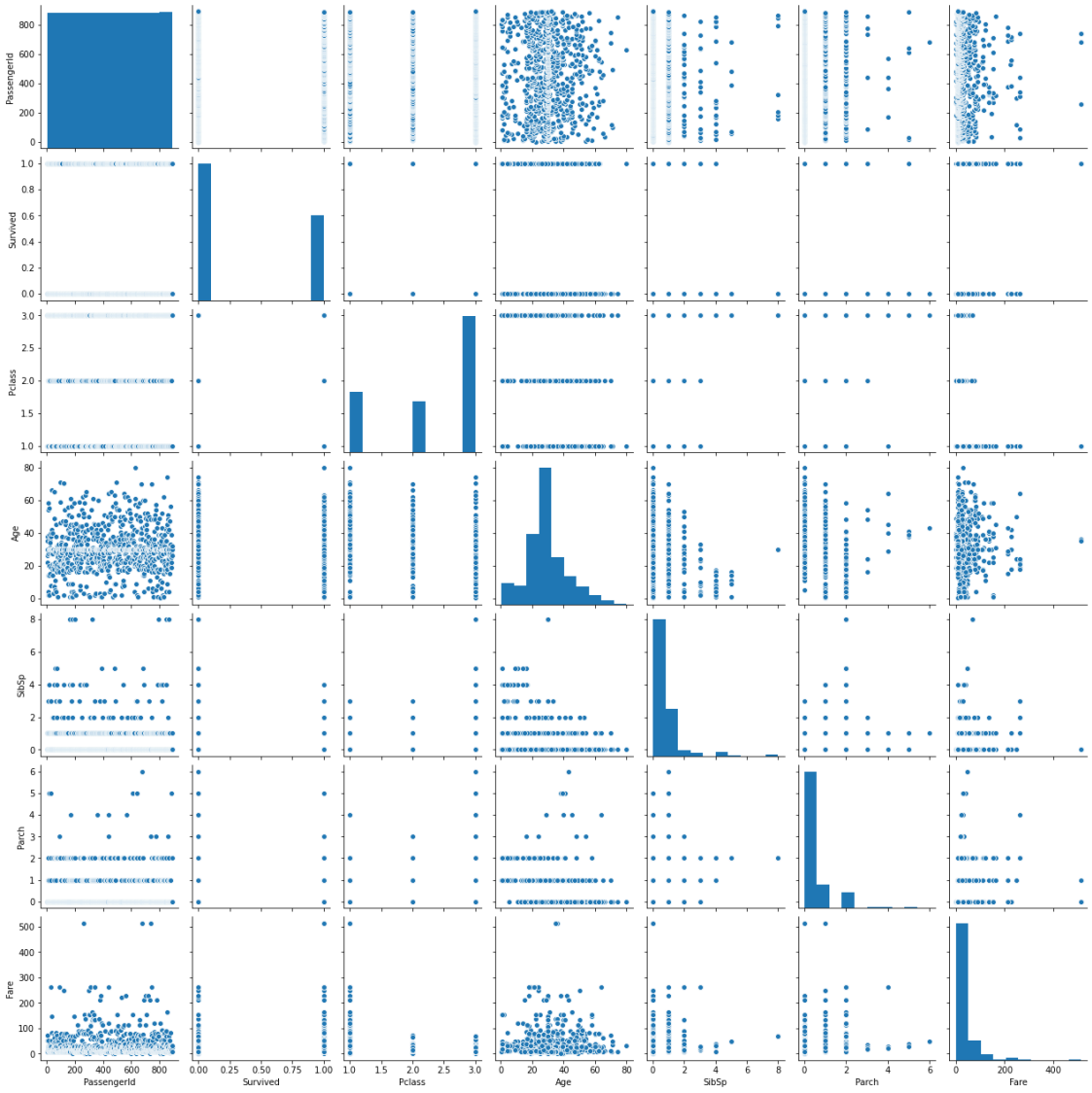
[22]:
```
sns.catplot(x= 'Sex', y = 'Age', data=data, kind = 'strip')
```

[22]: <seaborn.axisgrid.FacetGrid at 0x1b5c8d515e0>



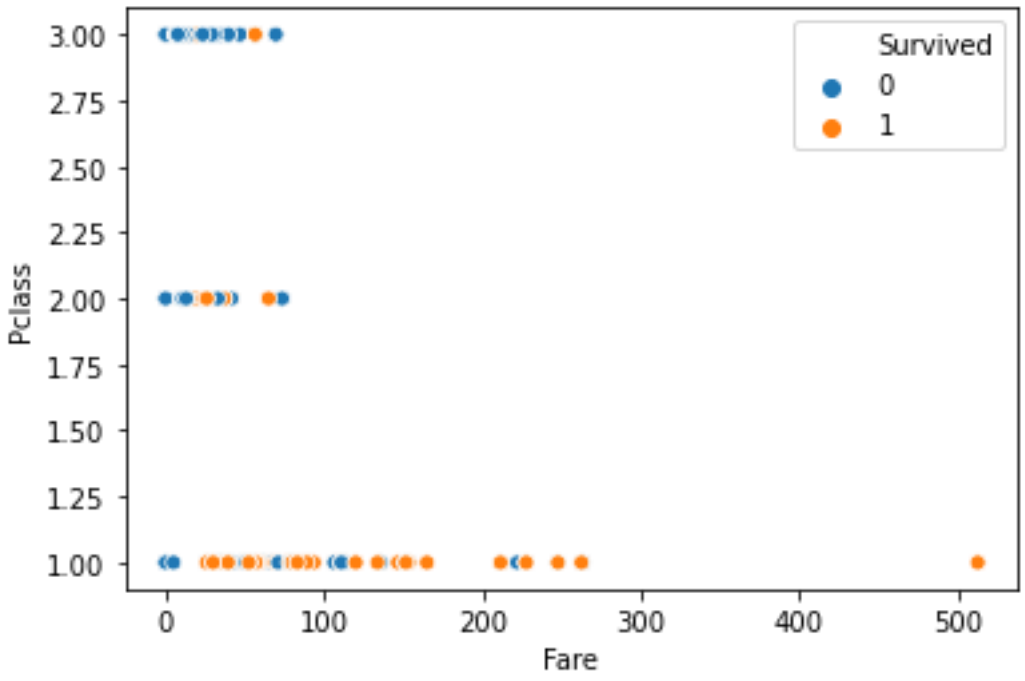
[23]:
```
sns.pairplot(data)
```

[23]: <seaborn.axisgrid.PairGrid at 0x1b5c8d9c490>



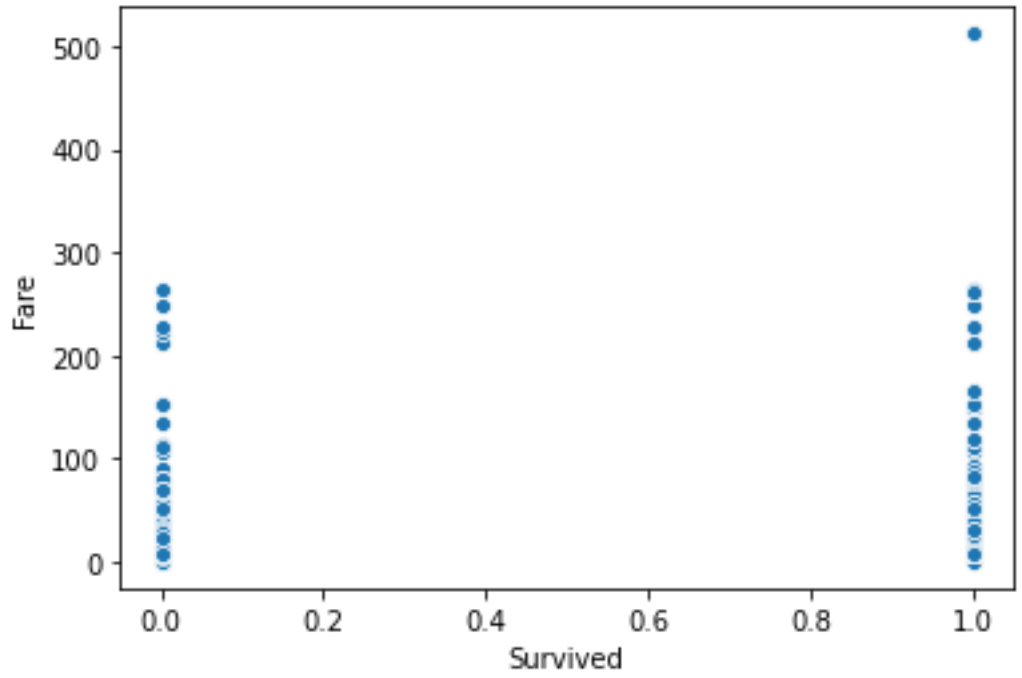
[24]:
```
sns.scatterplot(x = 'Fare', y = 'Pclass', hue = 'Survived',
data = data)
```

[24]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5cae10190>



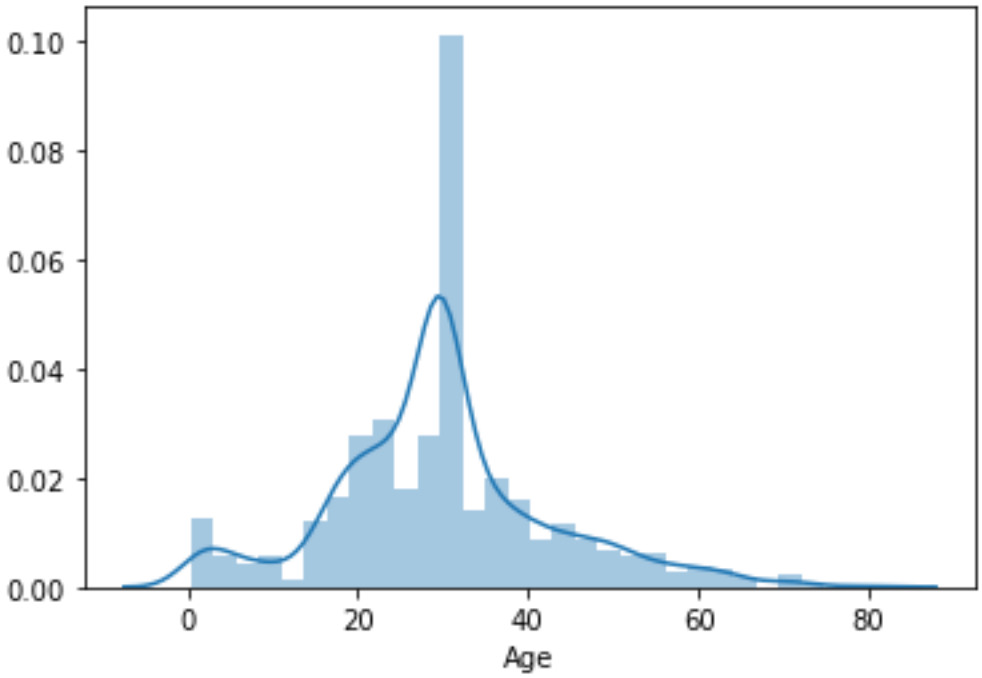
[25]:
```
sns.scatterplot(x = 'Survived', y = 'Fare', data = data)
```

[25]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5c8d84820>



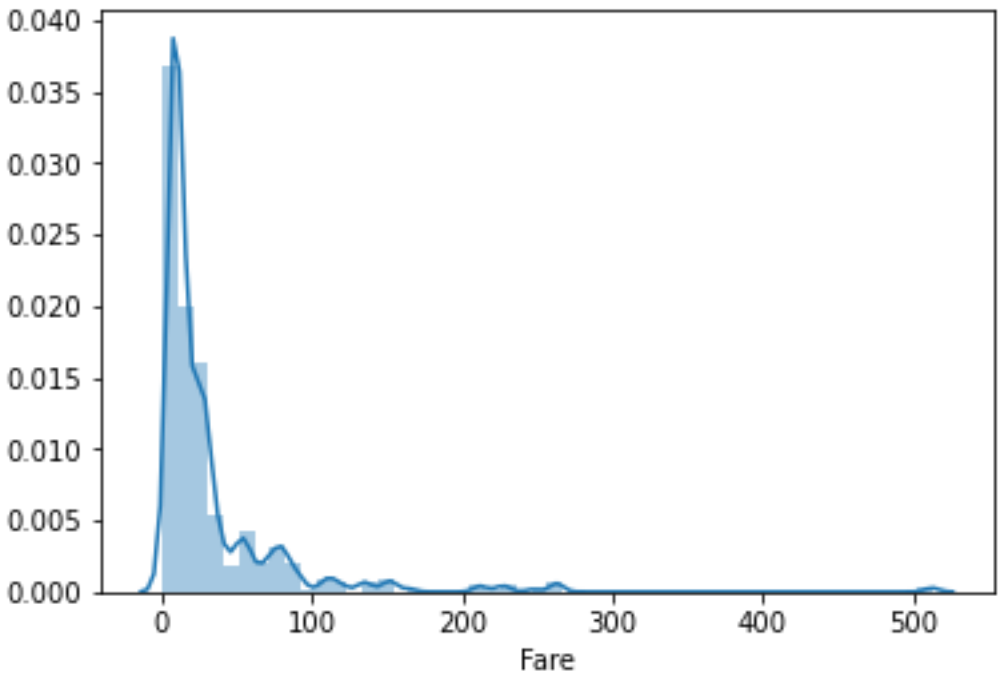
[26]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5cba89460>

[26]:
```
sns.distplot(data['Age'])
```



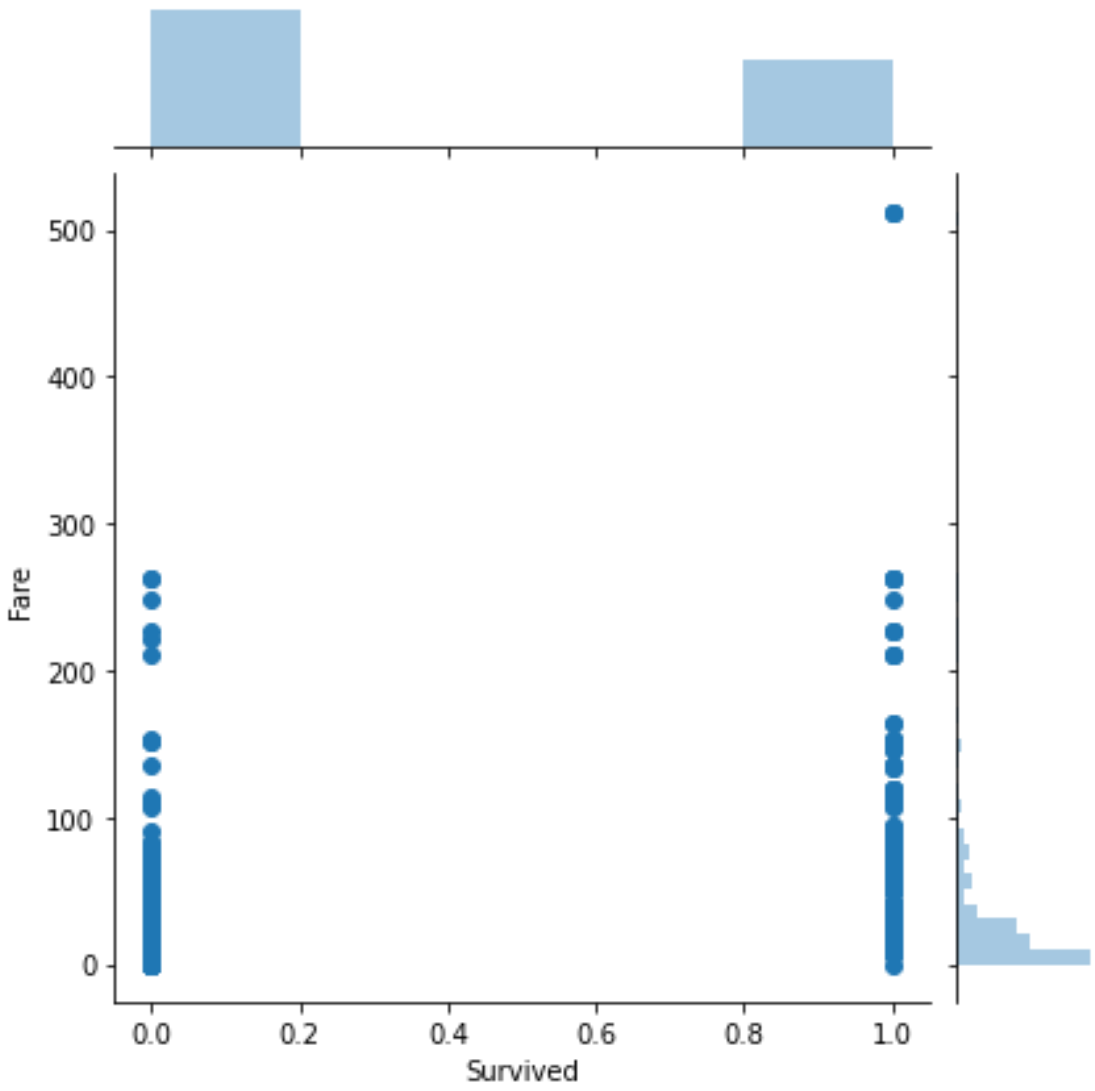
[27]:
```
sns.distplot(data['Fare'])
```

[27]: <matplotlib.axes._subplots.AxesSubplot at 0x1b5cbb34f40>



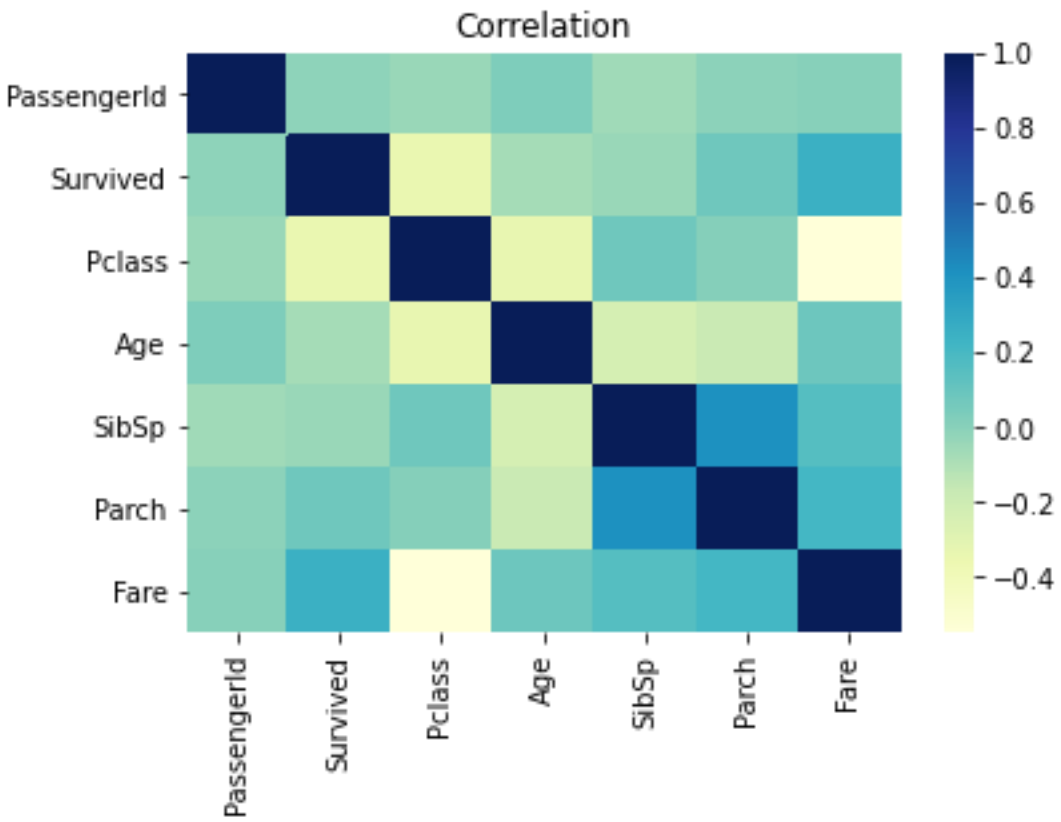
[28]:
```
sns.jointplot(x = "Survived", y = "Fare", kind = "scatter",
data = data)
```

[28]:
```
<seaborn.axisgrid.JointGrid at 0x1b5cbbf62b0>
```



[29]:
```
tc = data.corr()
sns.heatmap(tc, cmap="YlGnBu")
plt.title('Correlation')
```

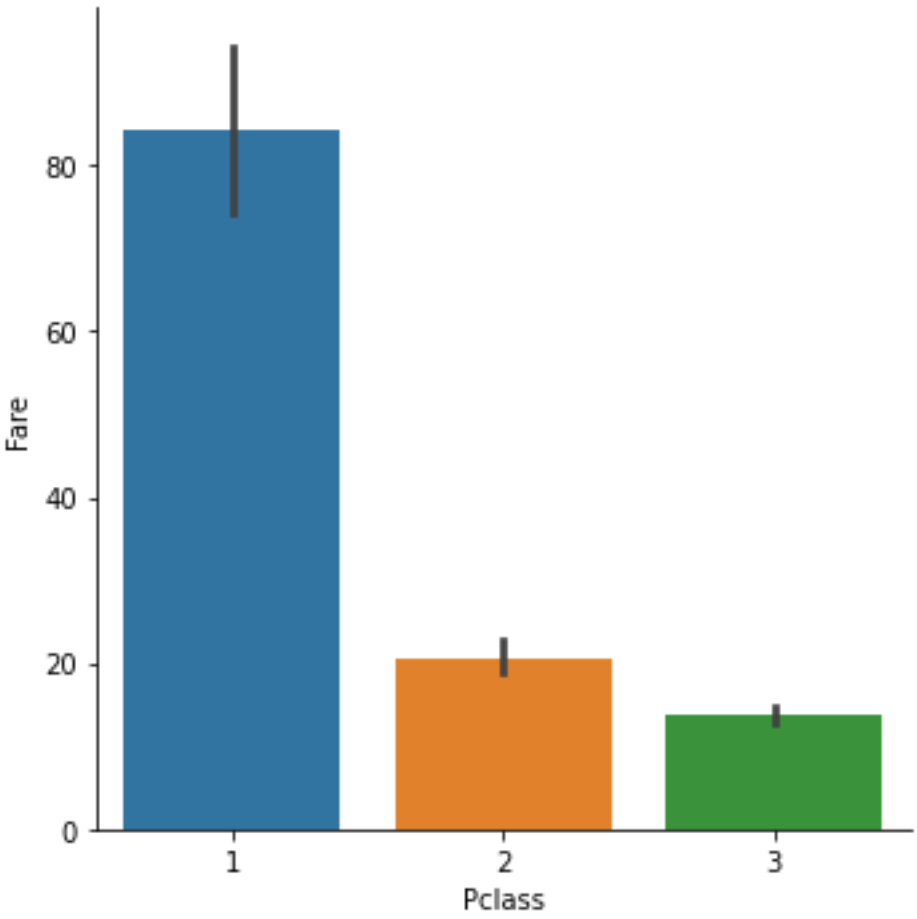
[29]: Text(0.5, 1.0, 'Correlation')



Price of Ticket for each passenger is distributed

[33]:
```
sns.catplot(x='Pclass', y='Fare', data=data, kind='bar')
```

[33]:
```
<seaborn.axisgrid.FacetGrid at 0x1b5ca67bcd0>
```


Fare
80
60
40
20
0
1
2
3
Pclass