

```
import pandas as pd
```

```
import numpy as np
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('Titanic.csv')
```

```
df.head(12)
```

	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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S

df.isnull().sum()

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

```
df['Age'] = df['Age'].fillna(df['Age'].mean())
```

```
df['Cabin'] = df['Cabin'].fillna(df['Cabin'].mode()[0])
```

```
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])
```

```
df.isnull().sum()
```

PassengerId 0

Survived 0

Pclass 0

Name 0

Sex 0

Age 0

SibSp 0

Parch 0

Ticket 0

Fare 0

Cabin 0

Embarked 0

dtype: int64

df.describe()

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	891.000000	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	13.002015	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	22.000000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	29.699118	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

Column Non-Null Count Dtype

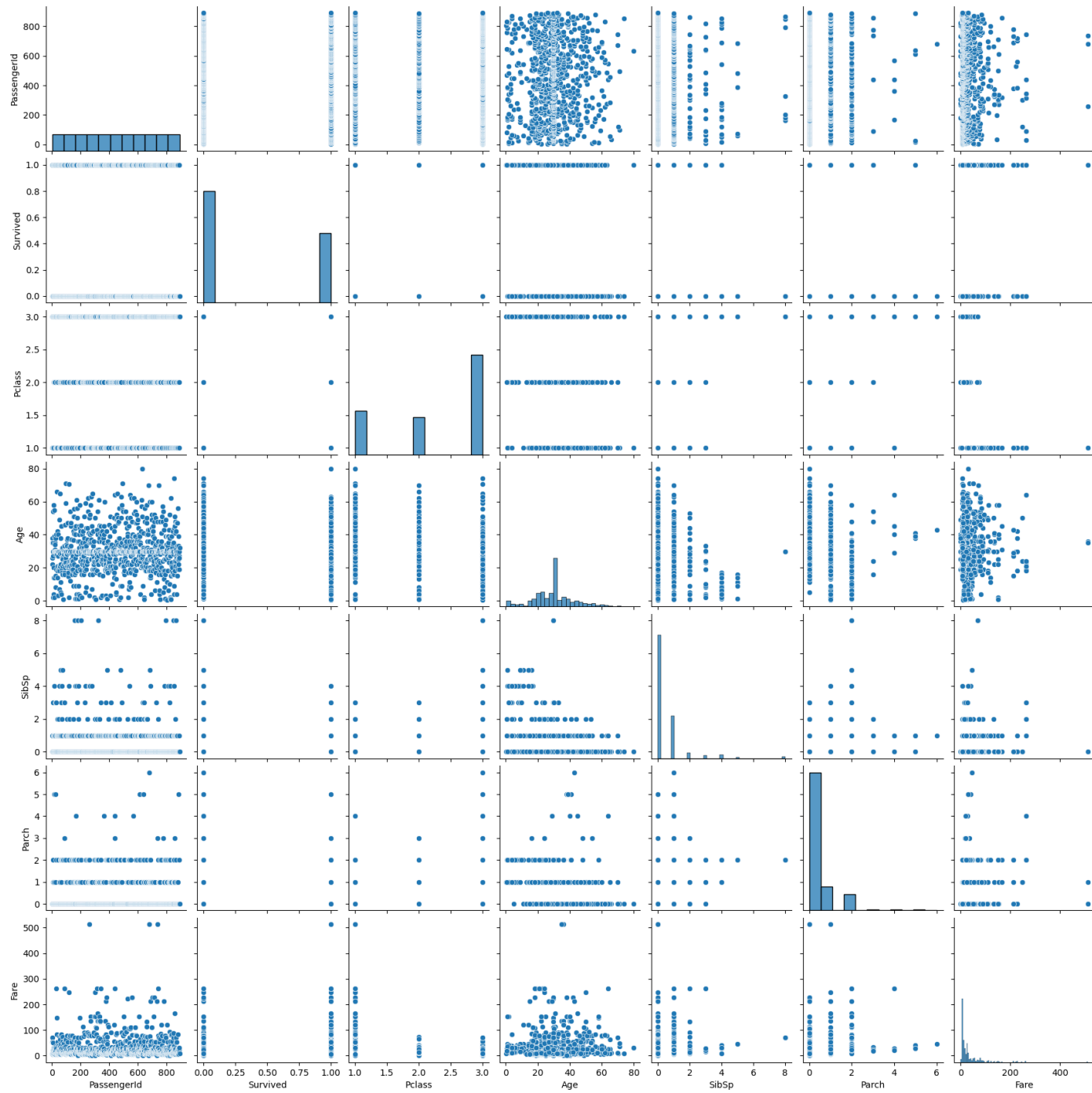
--- -----

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 891 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 891 non-null object
11 Embarked 891 non-null object

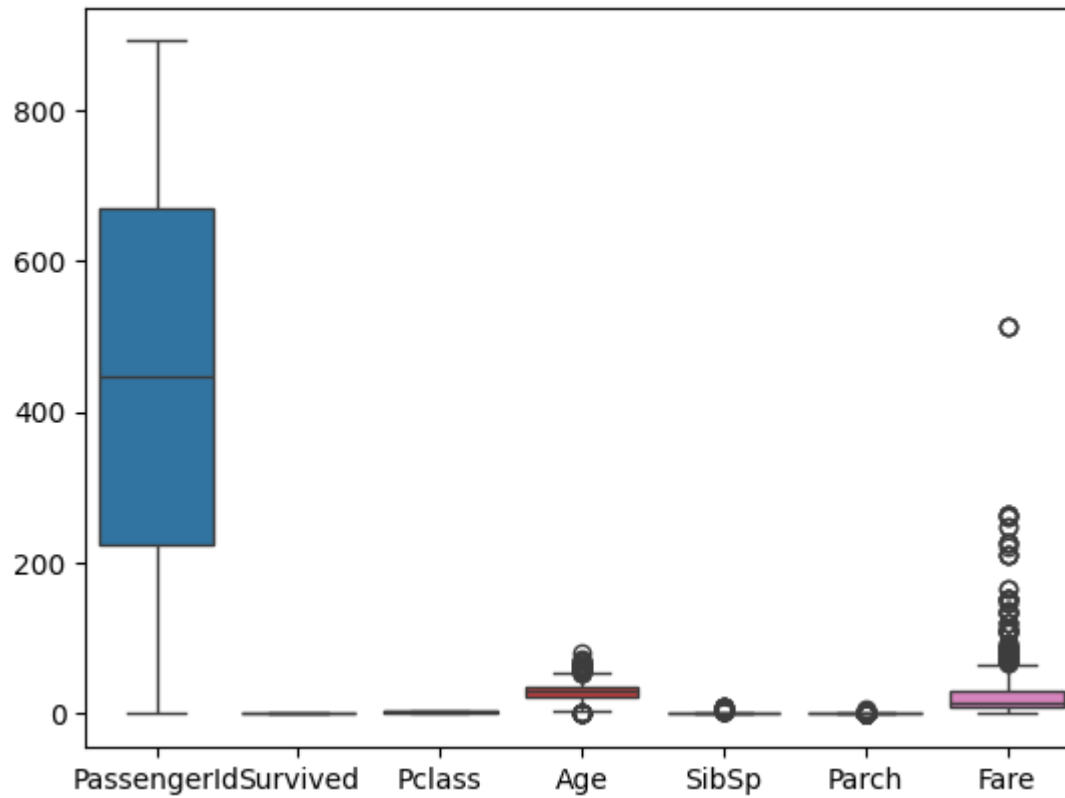
dtypes: float64(2), int64(5), object(5)

memory usage: 83.7+ KB

sns.pairplot(**data**=df)



```
sns.boxplot(data=df)
```



```
df['Sex'].value_counts()
```

Sex

male 577

female 314

Name: count, dtype: int64

```
df['Pclass'].value_counts()
```

Pclass

3 491

1 216

2 184

Name: count, dtype: int64

df['Survived'].value_counts()

Survived

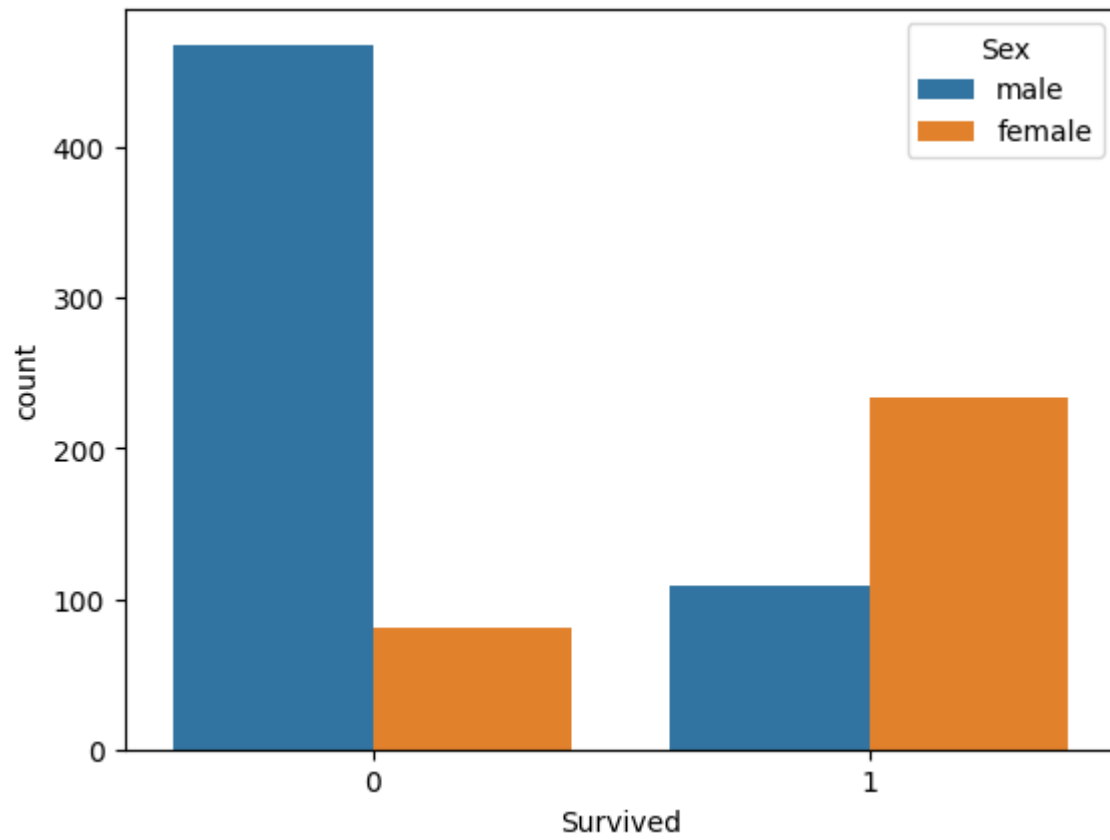
0 549

1 342

Name: count, dtype: int64

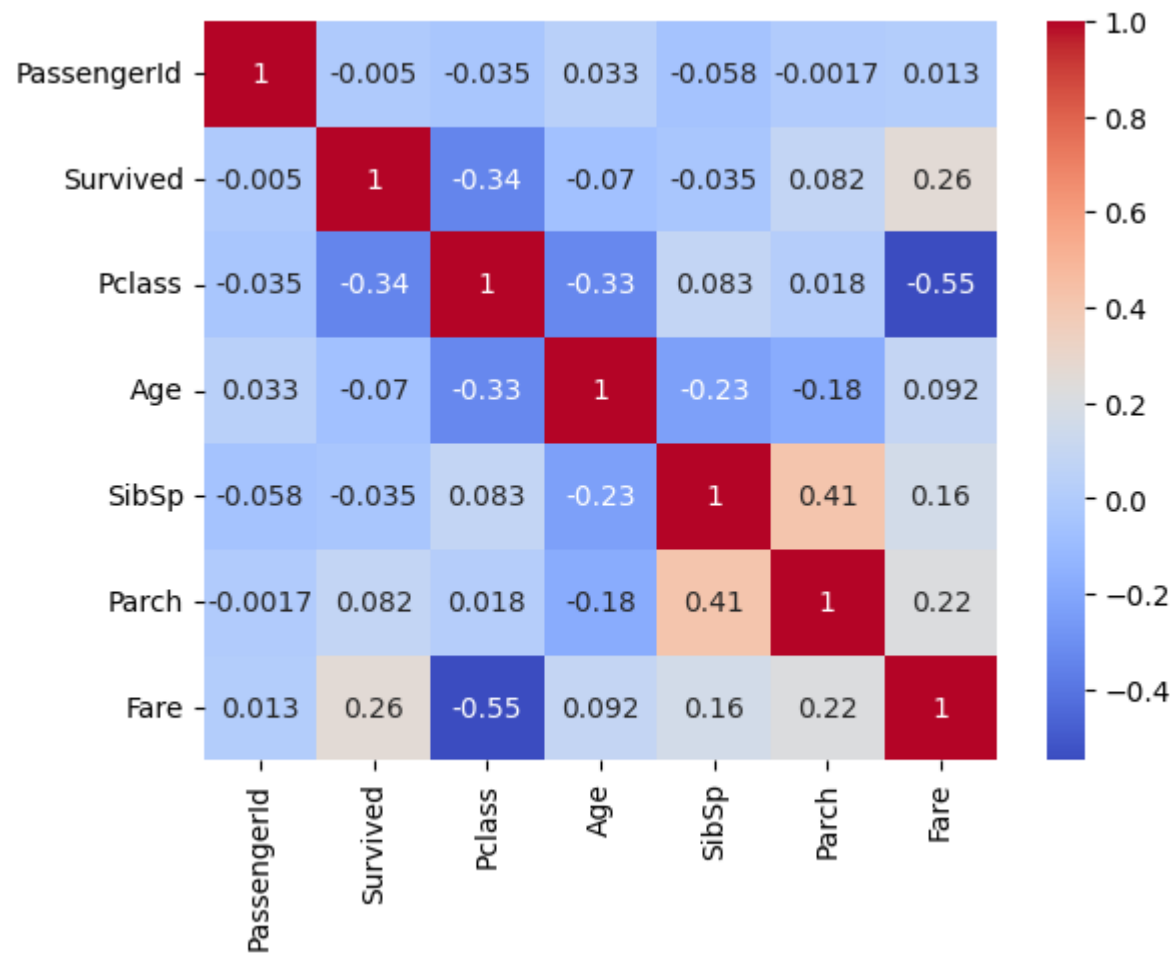
sns.countplot(x='Survived', hue='Sex', data=df)

plt.show()



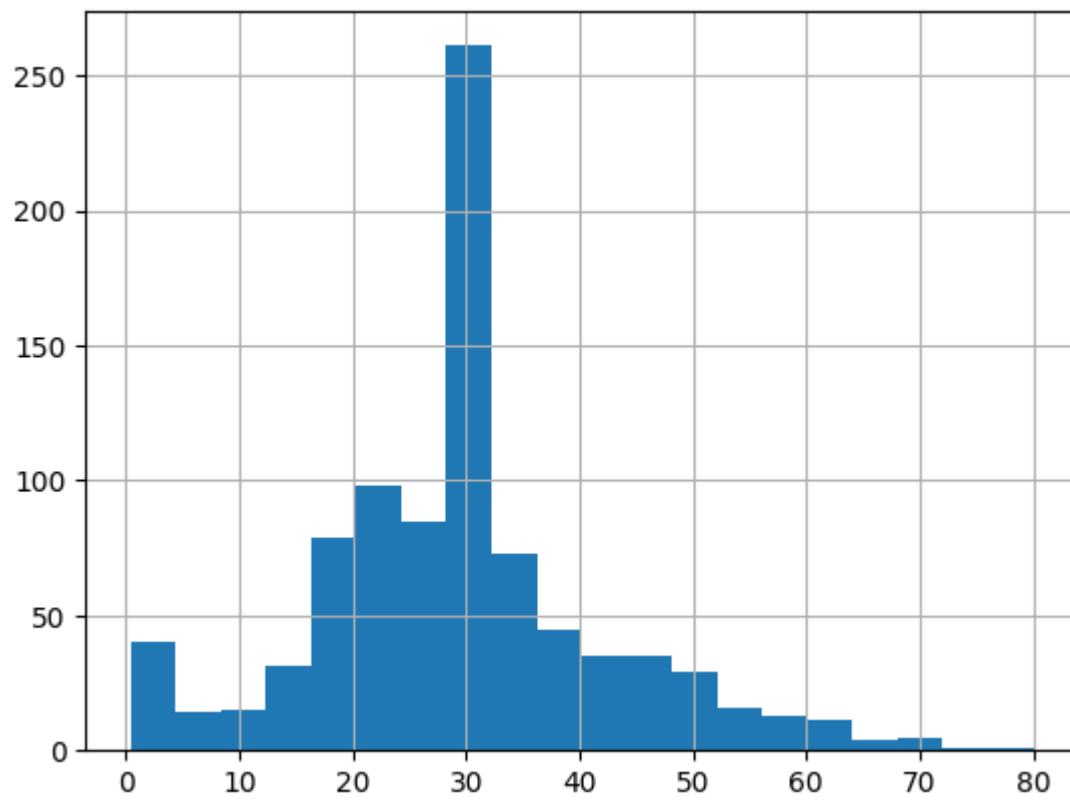
```
sns.heatmap(df.select_dtypes(include='number').corr(), annot=True, cmap='coolwarm')
```

```
plt.show()
```



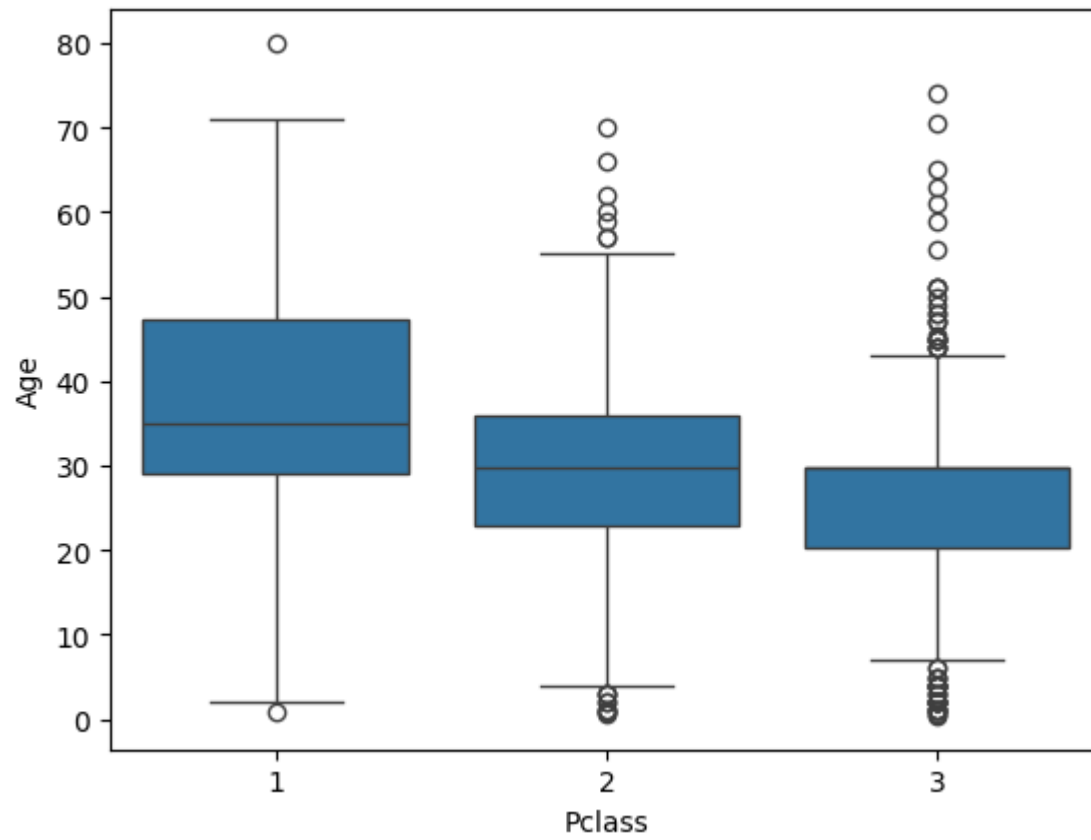
```
df['Age'].hist(bins=20)
```

```
plt.show()
```



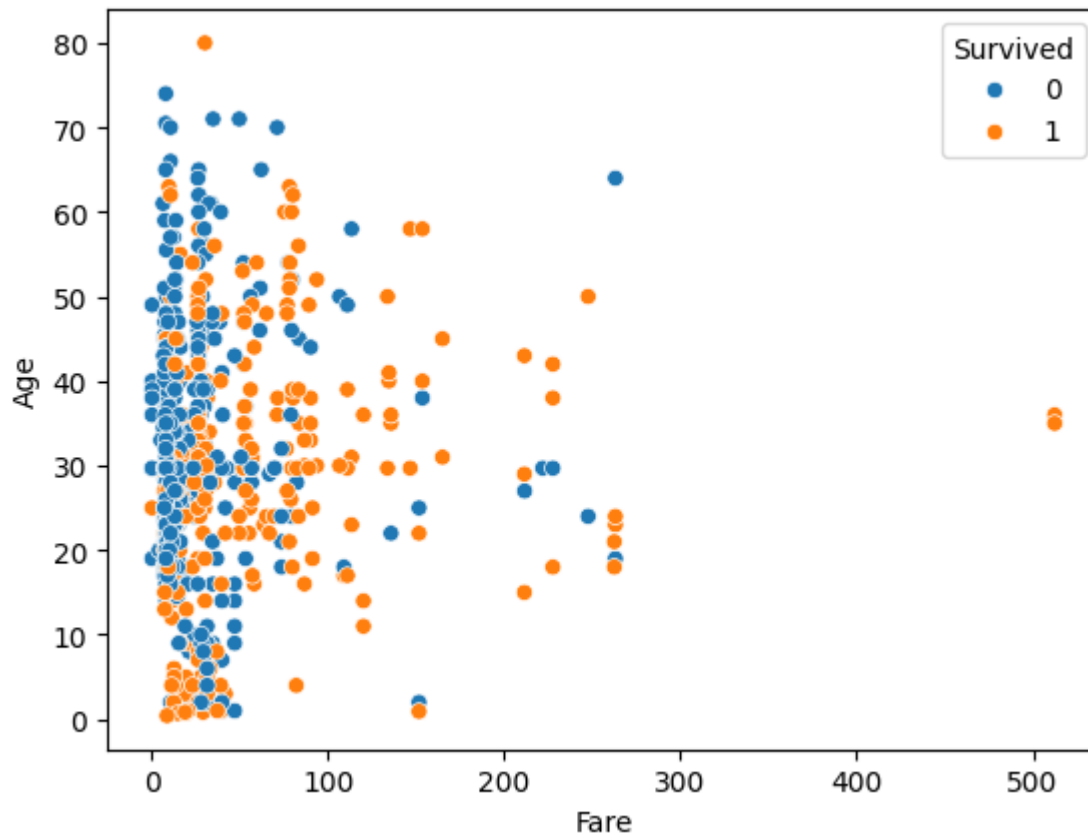
```
sns.boxplot(x='Pclass', y='Age', data=df)
```

```
plt.show()
```



```
sns.scatterplot(x='Fare', y='Age', hue='Survived', data=df)
```

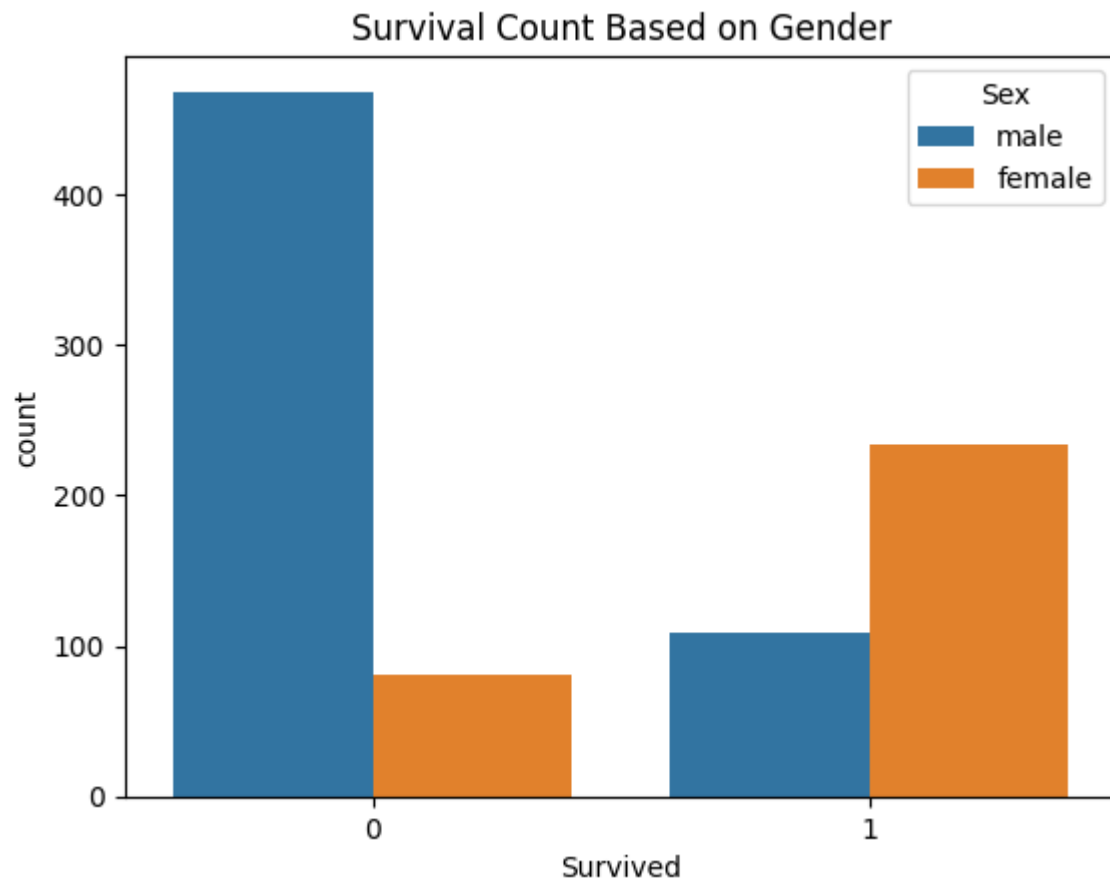
```
plt.show()
```



```
sns.countplot(x='Survived', hue='Sex', data=df)
```

```
plt.title('Survival Count Based on Gender')
```

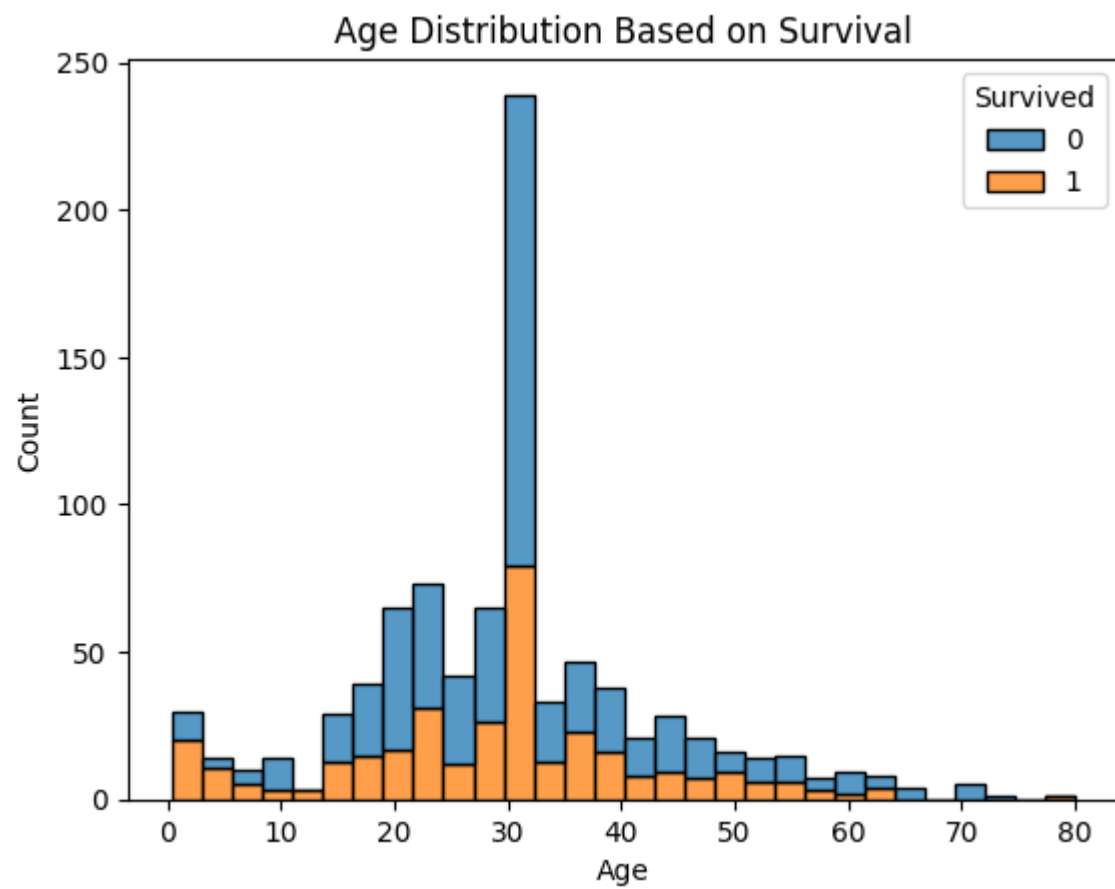
```
plt.show()
```



```
sns.histplot(data=df, x='Age', hue='Survived', multiple='stack')
```

```
plt.title('Age Distribution Based on Survival')
```

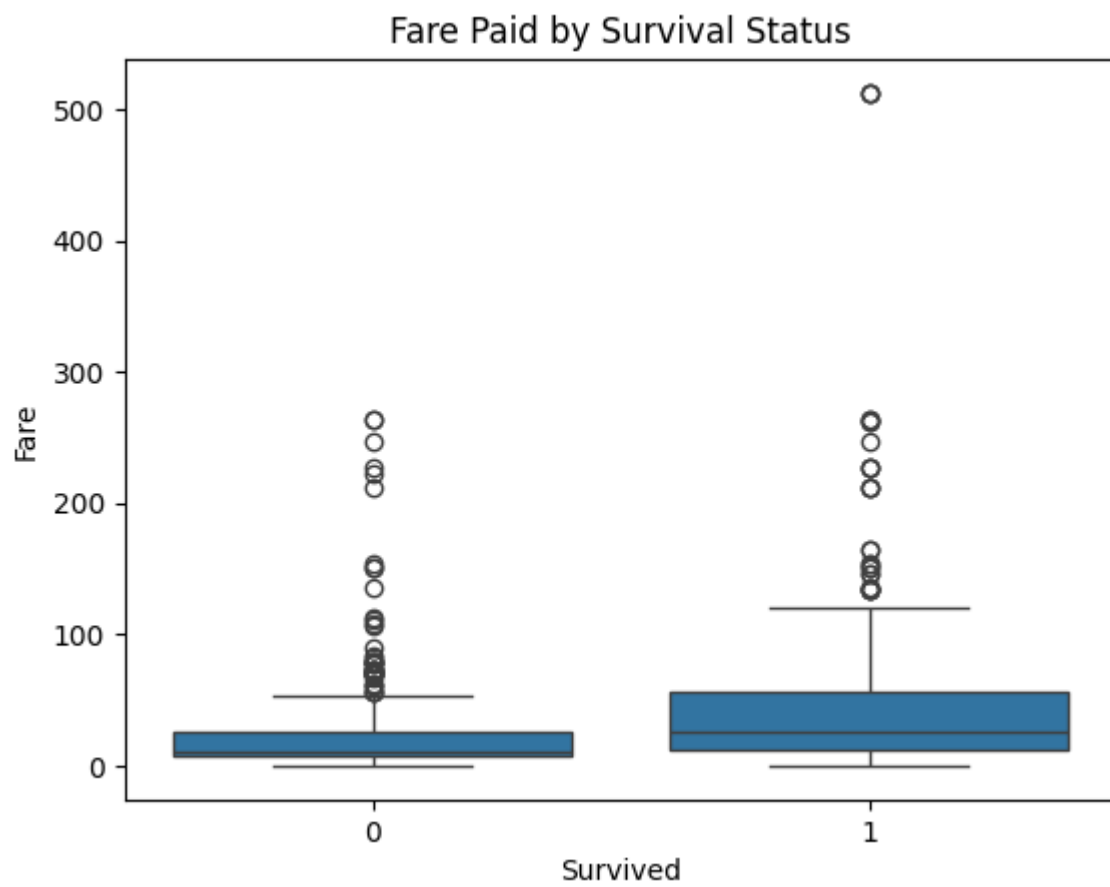
```
plt.show()
```



```
sns.boxplot(x='Survived', y='Fare', data=df)
```

```
plt.title('Fare Paid by Survival Status')
```

```
plt.show()
```



```
sns.countplot(x='Pclass', hue='Survived', data=df)
```

```
plt.title('Passenger Class vs Survival')
```

```
plt.show()
```


Passenger Class vs Survival

