In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings(action='ignore')
```

In [2]:
```
df=pd.read_csv(r"C:\Users\Shree\Downloads\titanic.csv")
df
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

In [3]:
```
df.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0

In [4]:
```
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 [5]:
```
df.shape
```

Out[5]: (891, 12)

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

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	False	False	False	False	False	False	False	False	False	False	T
1	False	False	False	False	False	False	False	False	False	False	Fa
2	False	False	False	False	False	False	False	False	False	False	T
3	False	False	False	False	False	False	False	False	False	False	Fa
4	False	False	False	False	False	False	False	False	False	False	T
...	...	...	...	...	...	...	...	...	...	...	
886	False	False	False	False	False	False	False	False	False	False	T
887	False	False	False	False	False	False	False	False	False	False	Fa
888	False	False	False	False	False	True	False	False	False	False	T
889	False	False	False	False	False	False	False	False	False	False	Fa
890	False	False	False	False	False	False	False	False	False	False	T

891 rows × 12 columns

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

Out[8]:
```
866
```

In [9]:
```
df.dtypes
```

Out[9]:
```
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
```

In [10]:
```
df.drop(['Cabin'],axis=1,inplace=True)
```

In [11]:
```
df.isnull().sum()
```

Out[11]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         2
dtype: int64
```

In [12]:
```
df['Age'].fillna(df['Age'].mean(),inplace=True)
df
```

Out[12]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A 211
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 175
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O 31012
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	1138
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	3734
...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	2115
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	1120
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W./ 66
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	1113
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	3703

891 rows × 11 columns

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64
```

In [14]:
```
df['Embarked'].value_counts()#count of each unique value
```

Out[14]:
```
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64
```

In [15]:
```
df['Embarked'].fillna('S',inplace=True)
```

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [17]:
```
df.describe()
```

Out[17]:

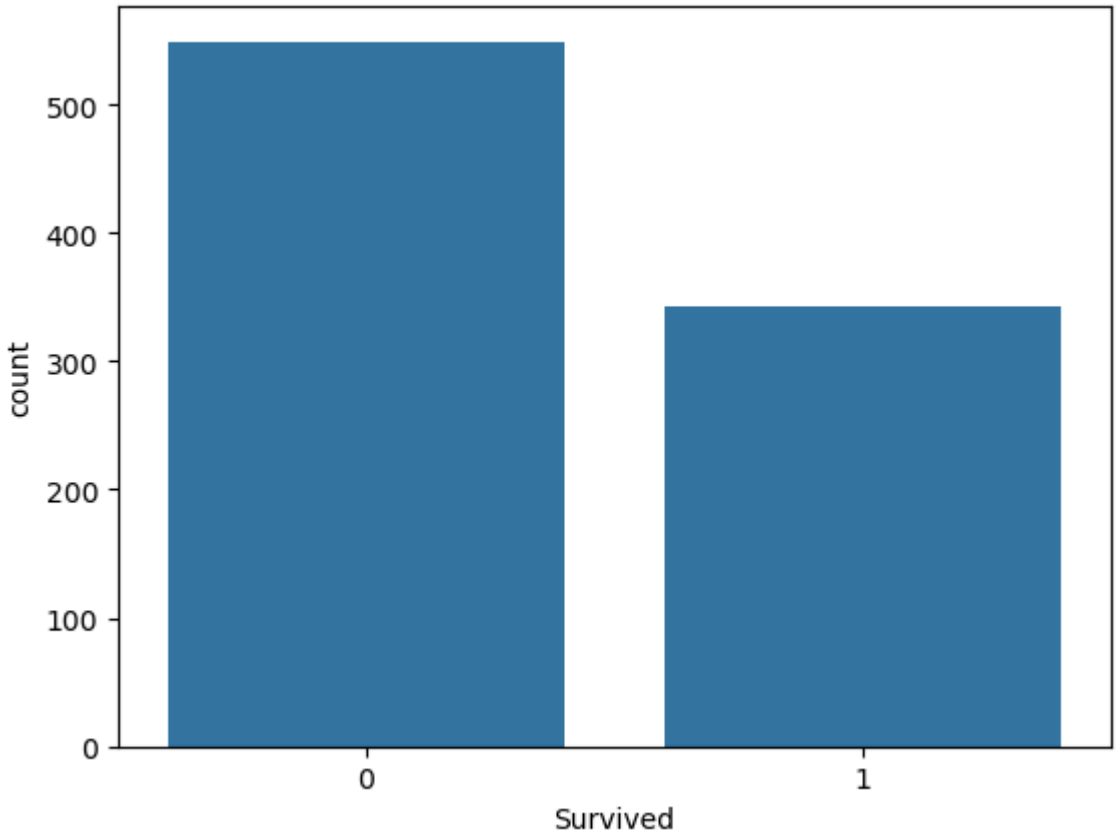
	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204
std	257.353842	0.486592	0.836071	13.002015	1.102743	0.806057	49.693
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.910
50%	446.000000	0.000000	3.000000	29.699118	0.000000	0.000000	14.454
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329

In [18]:
```
df['Survived'].value_counts()
```

Out[18]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

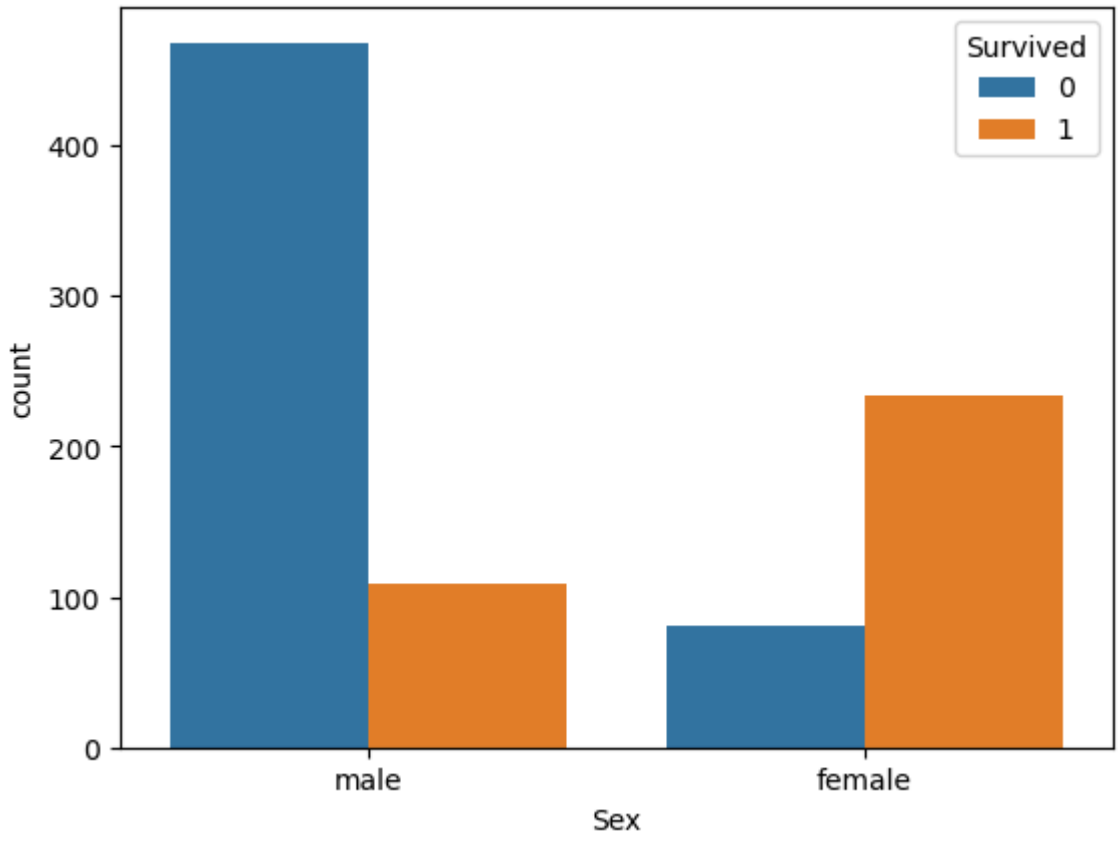
In [19]:
```
sns.countplot(x='Survived',data=df)
```

Out[19]:
```
<Axes: xlabel='Survived', ylabel='count'>
```



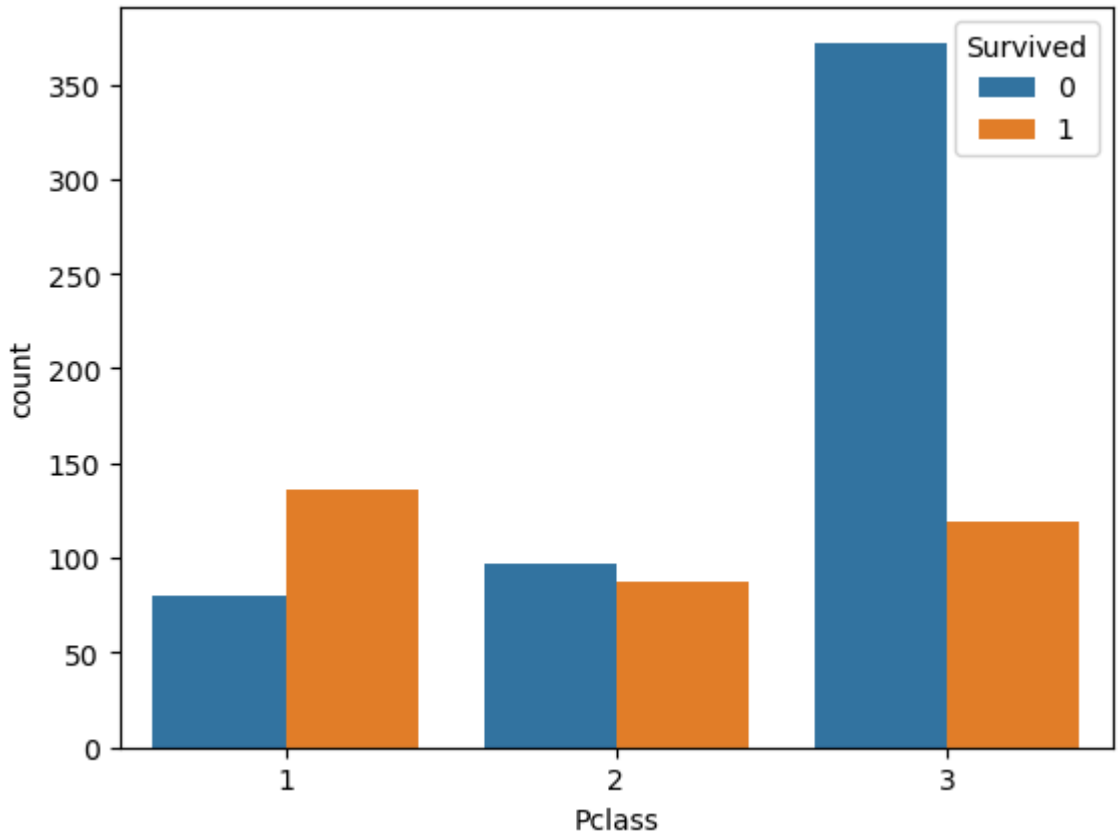
In [20]:
```
sns.countplot(x='Sex',hue='Survived',data=df)
```

Out[20]:
```
<Axes: xlabel='Sex', ylabel='count'>
```



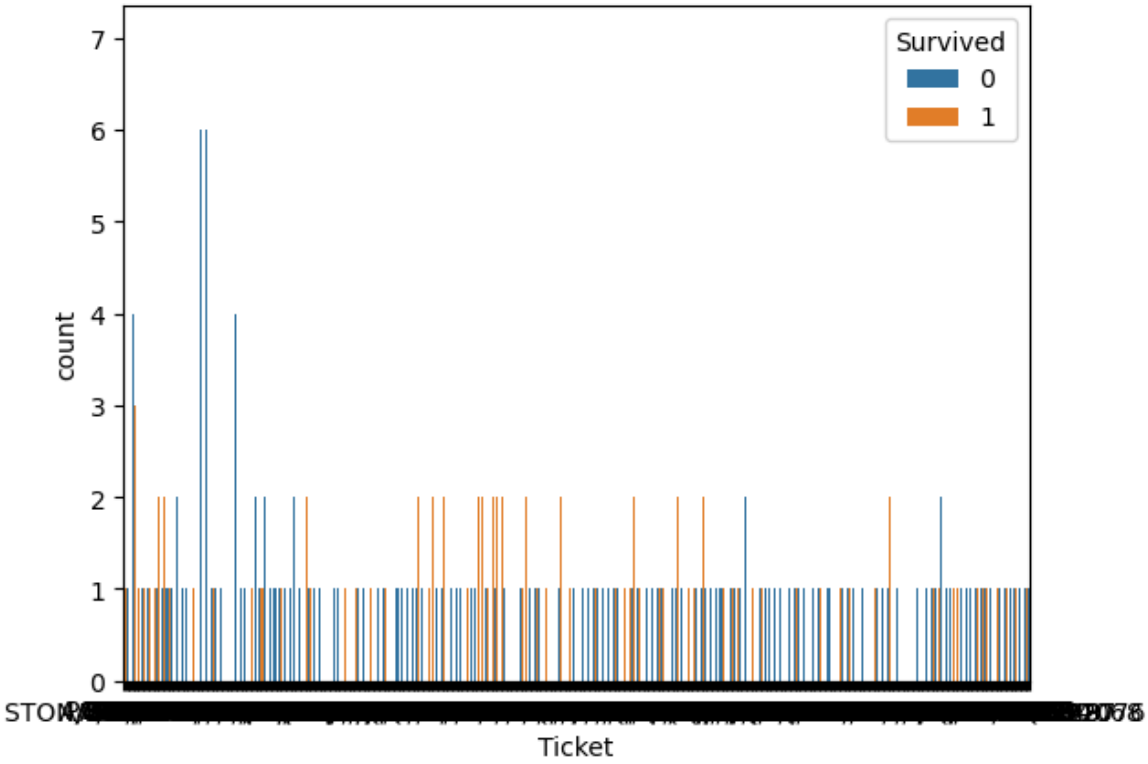
In [21]:
```
sns.countplot(x='Pclass',hue='Survived',data=df)
```

Out[21]: <Axes: xlabel='Pclass', ylabel='count'>



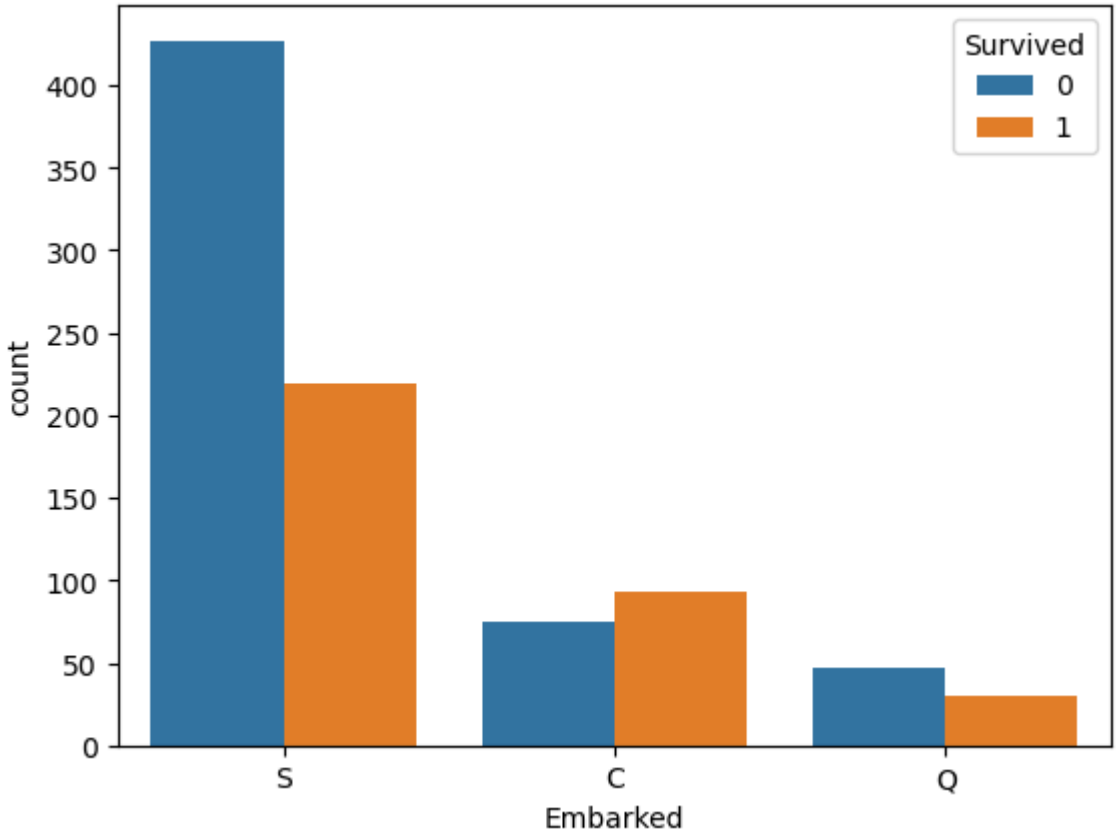
In [22]:
```
sns.countplot(x='Ticket',hue='Survived',data=df)
```

Out[22]: <Axes: xlabel='Ticket', ylabel='count'>



In [23]:
```
sns.countplot(x='Embarked',hue='Survived',data=df)
```

Out[23]: <Axes: xlabel='Embarked', ylabel='count'>

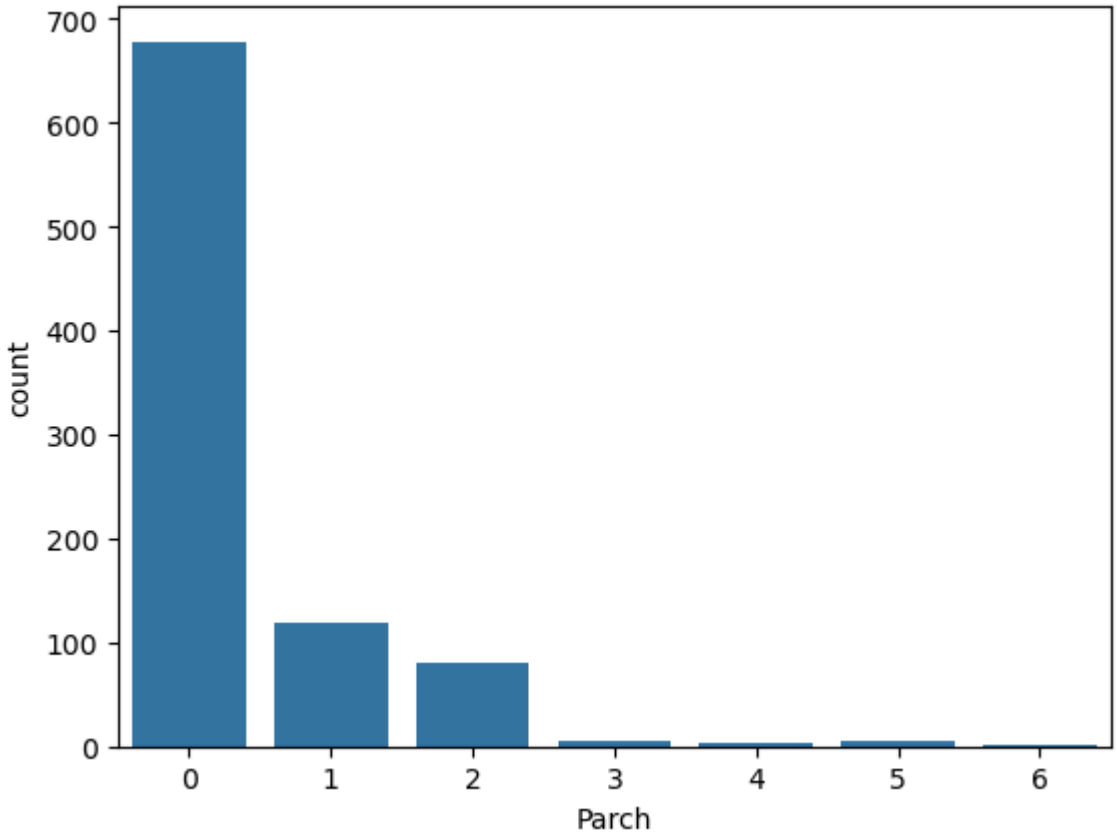


In [24]:
```
df['Parch'].value_counts()
```

Out[24]:
```
Parch
0    678
1    118
2     80
5      5
3      5
4      4
6      1
Name: count, dtype: int64
```

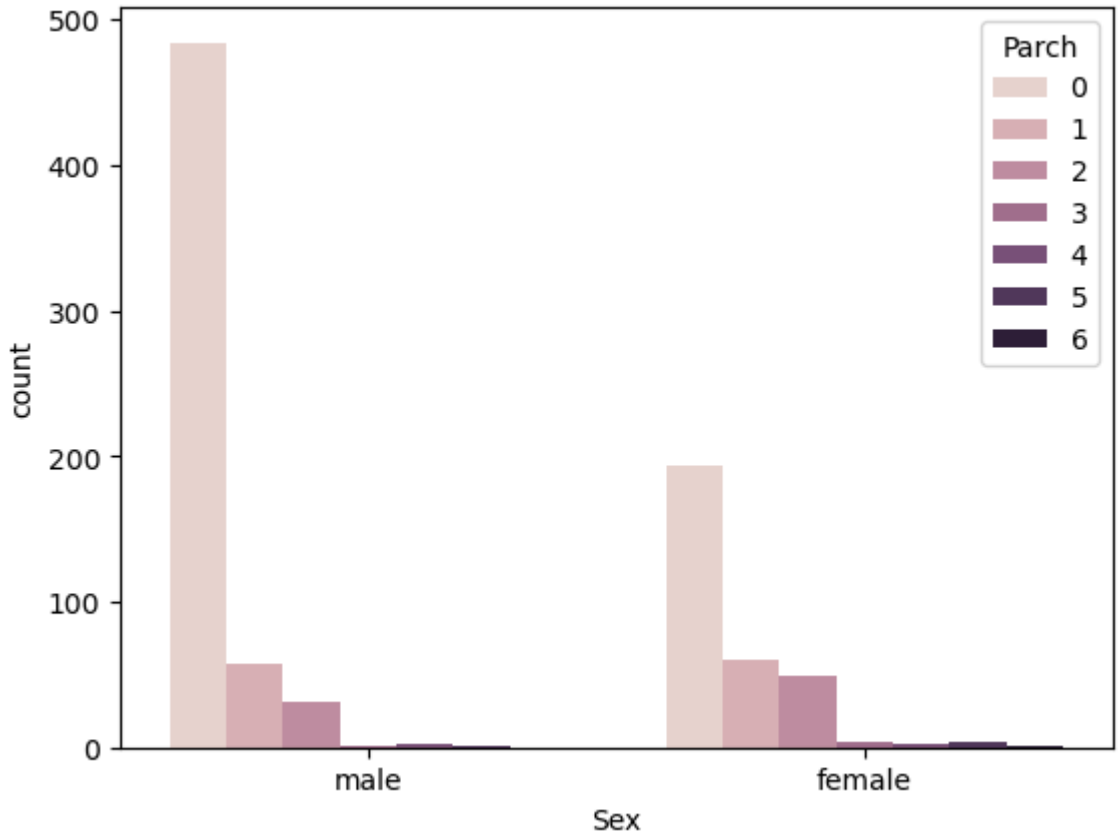
In [25]:
```
sns.countplot(x='Parch',data=df)
```

Out[25]:
```
<Axes: xlabel='Parch', ylabel='count'>
```



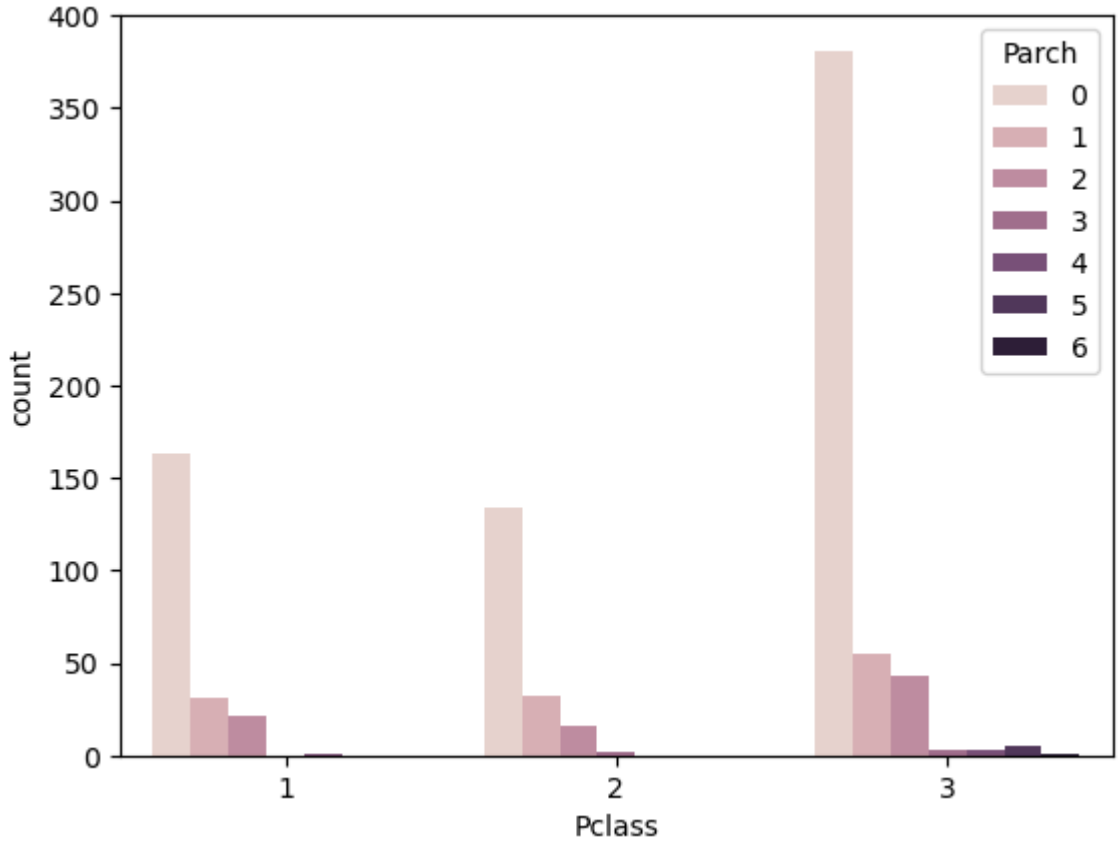
In [26]:
```
sns.countplot(x='Sex',hue='Parch',data=df)
```

Out[26]:
```
<Axes: xlabel='Sex', ylabel='count'>
```



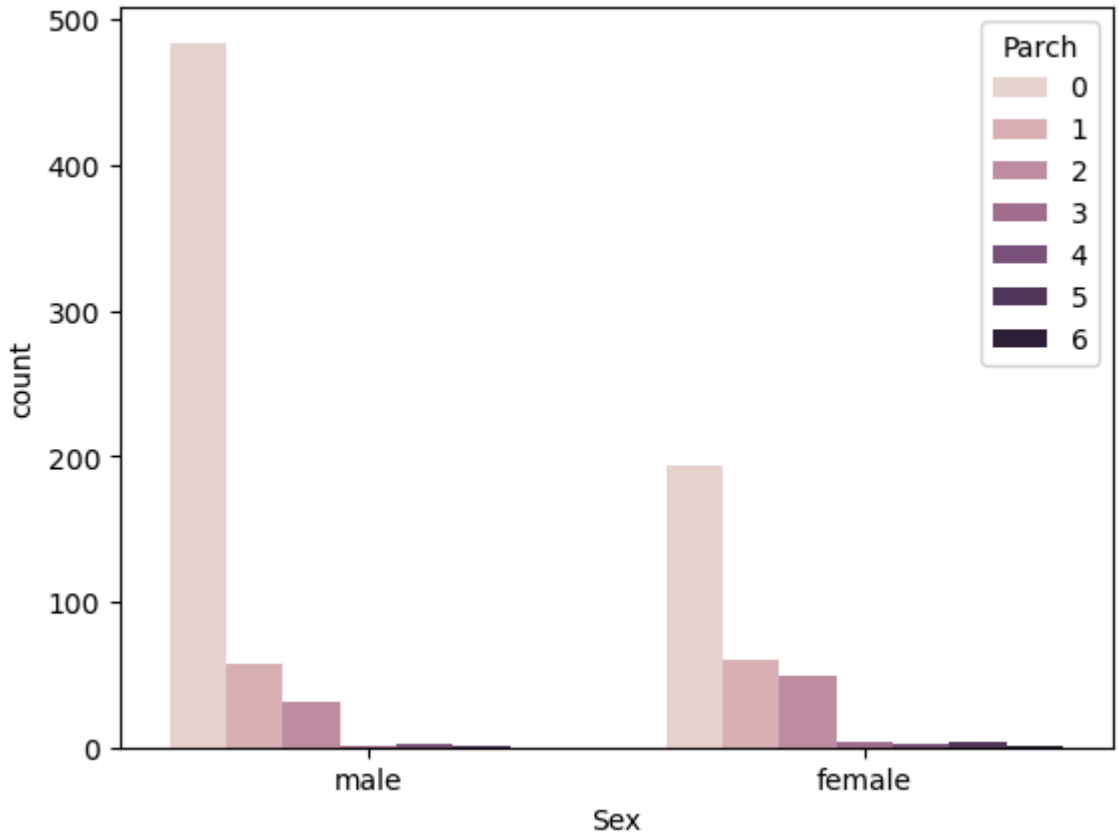
In [27]:
```
sns.countplot(x='Pclass',hue='Parch',data=df)
```

Out[27]: <Axes: xlabel='Pclass', ylabel='count'>



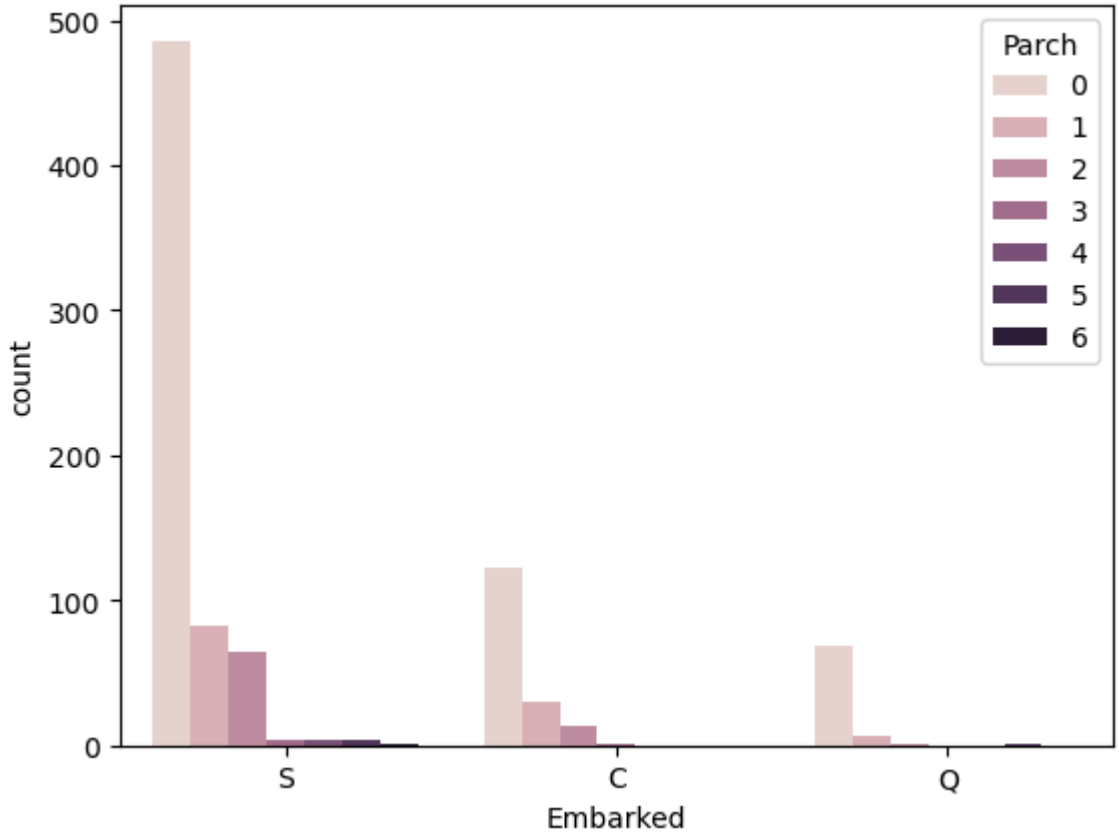
In [28]:
```
sns.countplot(x='Sex',hue='Parch',data=df)
```

Out[28]: <Axes: xlabel='Sex', ylabel='count'>



In [29]:
```
sns.countplot(x='Embarked',hue='Parch',data=df)
```

Out[29]: <Axes: xlabel='Embarked', ylabel='count'>



In [30]:
```
df['Survived'].value_counts()
```

Out[30]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

In [31]:
```
df['Pclass'].value_counts()
```

Out[31]:
```
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
```

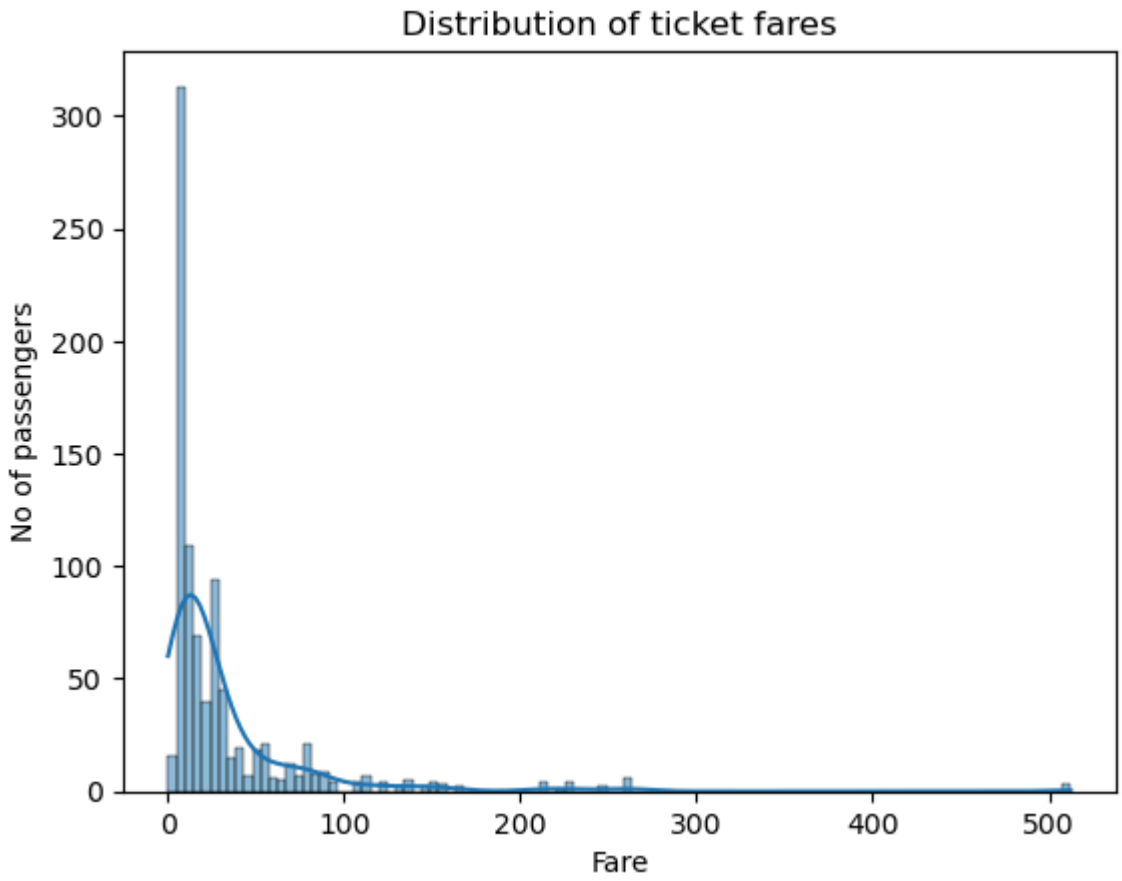
In [32]:
```
df['Fare'].value_counts()
```

Out[32]:
```
Fare
8.0500     43
13.0000    42
7.8958     38
7.7500     34
26.0000    31
           ..
35.0000     1
28.5000     1
6.2375      1
14.0000     1
10.5167     1
Name: count, Length: 248, dtype: int64
```

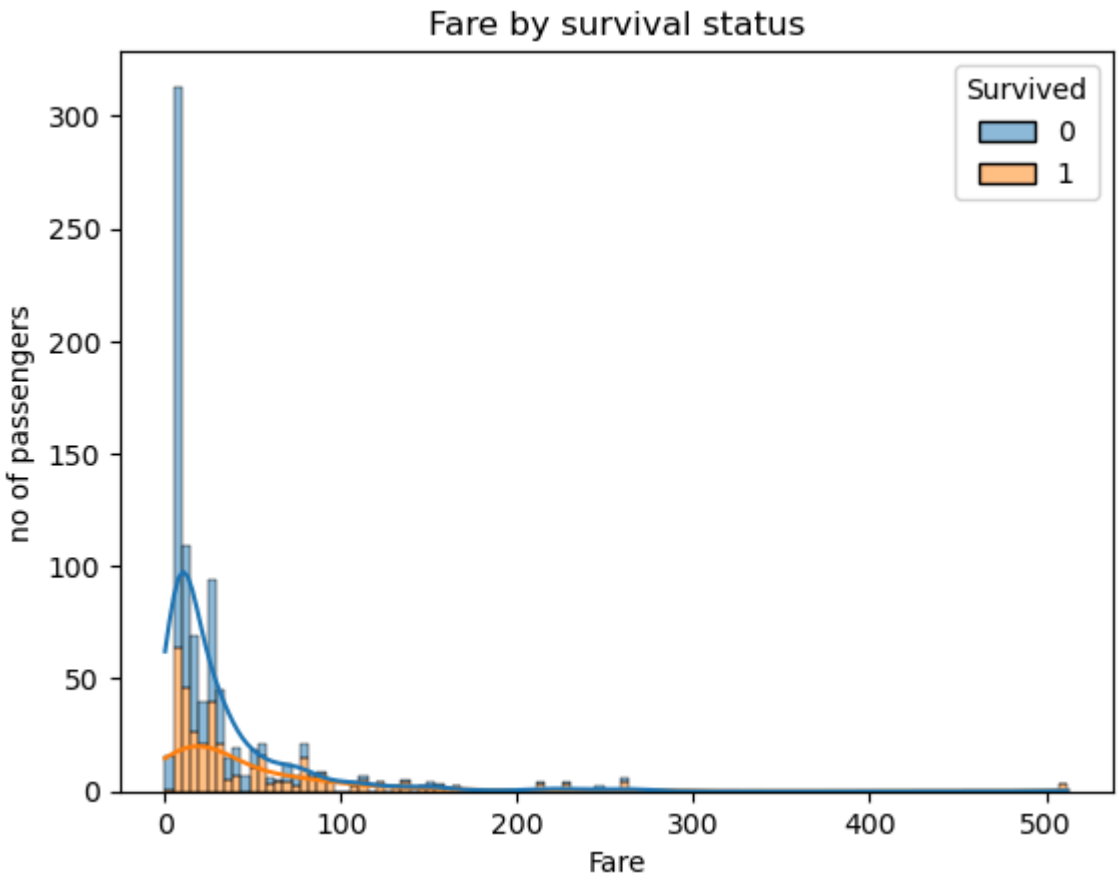
In [33]:
```
df['Parch'].value_counts()
```

Out[33]:
```
Parch
0    678
1    118
2     80
5      5
3      5
4      4
6      1
Name: count, dtype: int64
```

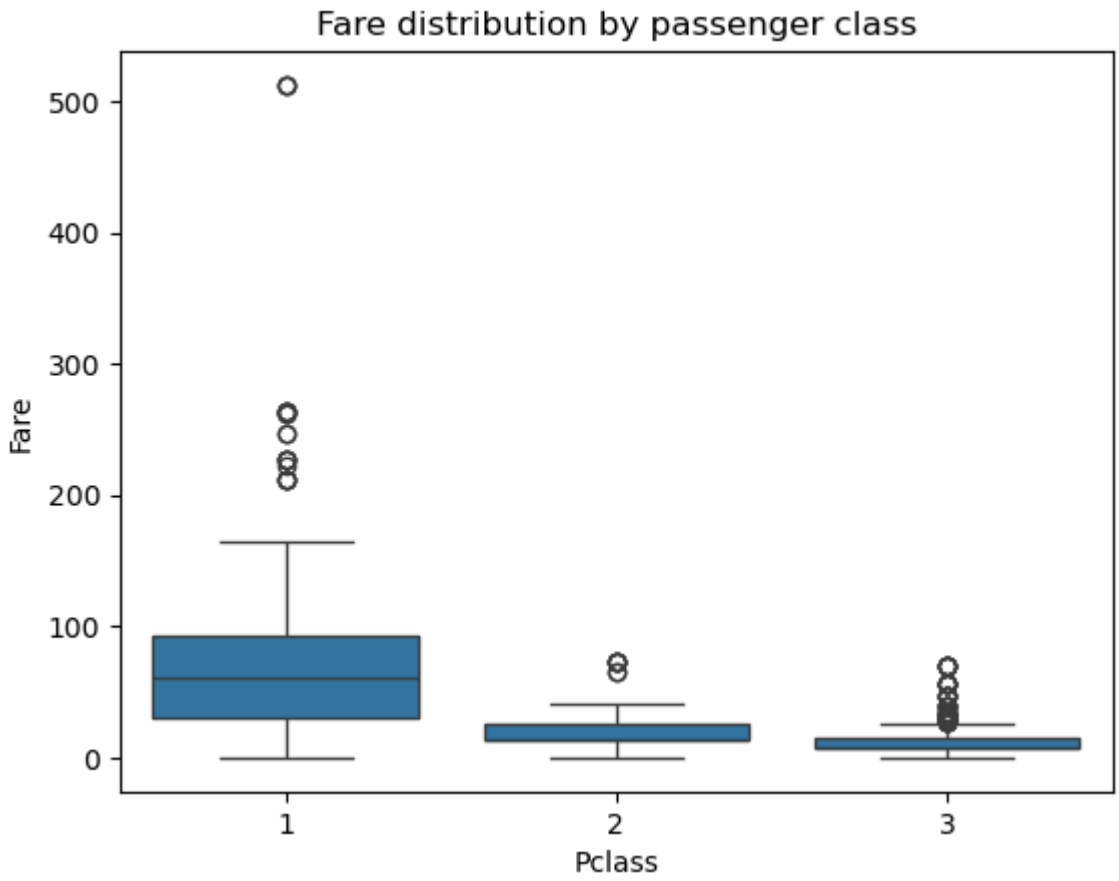
In [34]:
```
import matplotlib.pyplot as plt
sns.histplot(data=df,x='Fare',kde=True)
plt.title('Distribution of ticket fares')
plt.xlabel('Fare')
plt.ylabel('No of passengers')
plt.show()
```



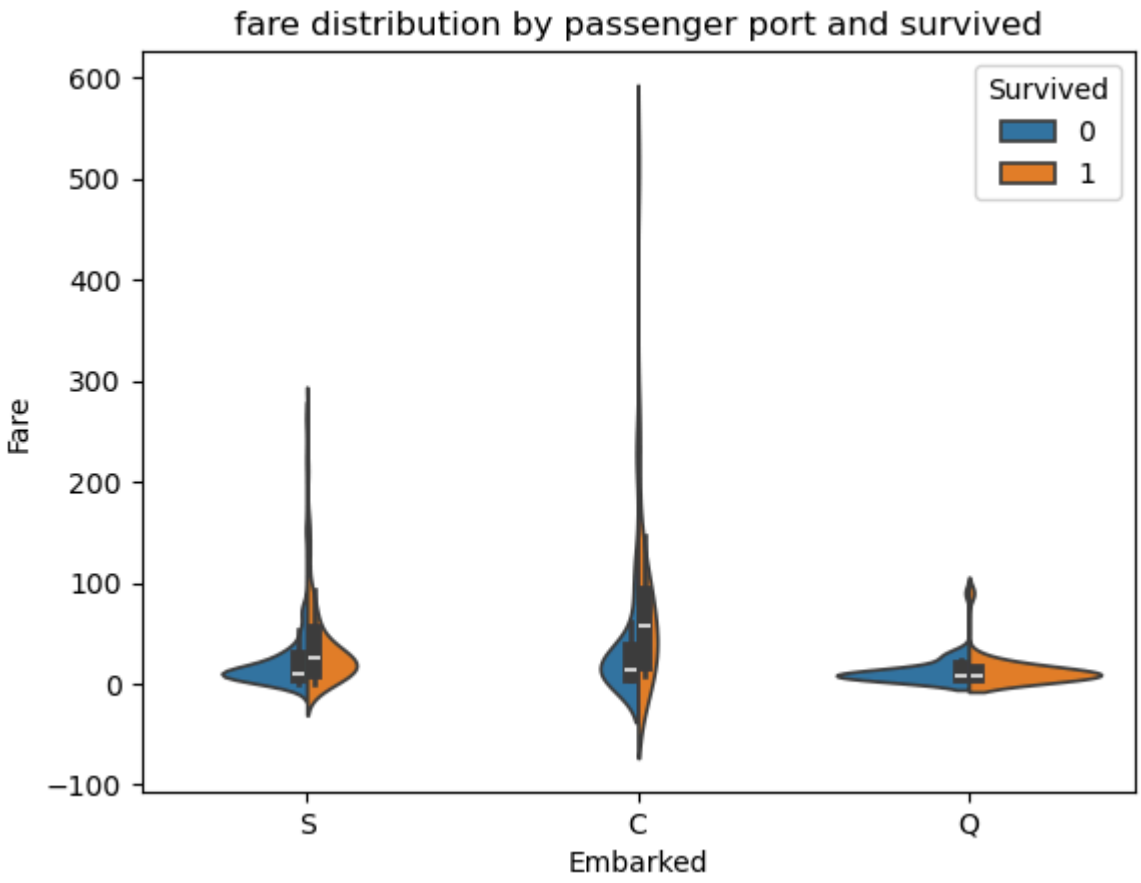
In [35]:
```
sns.histplot(data=df,x='Fare',hue='Survived',kde=True,multiple='stack')
plt.title('Fare by survival status')
plt.xlabel('Fare')
plt.ylabel('no of passengers')
plt.show()
```



In [36]:
```
sns.boxplot(data=df,x='Pclass',y='Fare')
plt.title('Fare distribution by passenger class')
plt.xlabel('Pclass')
plt.ylabel('Fare')
plt.show()
```



In [40]:
```
sns.violinplot(data=df,x='Embarked',y='Fare',hue='Survived',split=True)
plt.title('fare distribution by passenger port and survived')
plt.xlabel('Embarked')
plt.ylabel('Fare')
plt.show()
```


fare distribution by passenger port and survived
Survived
0
1
600
500
400
300
200
100
0
−100
Fare
S
C
Q
Embarked

In [ ]: