In [7]:
```
import pandas as pd
import matplotlib.pyplot as plt
import sns

_import seaborn as sns
```

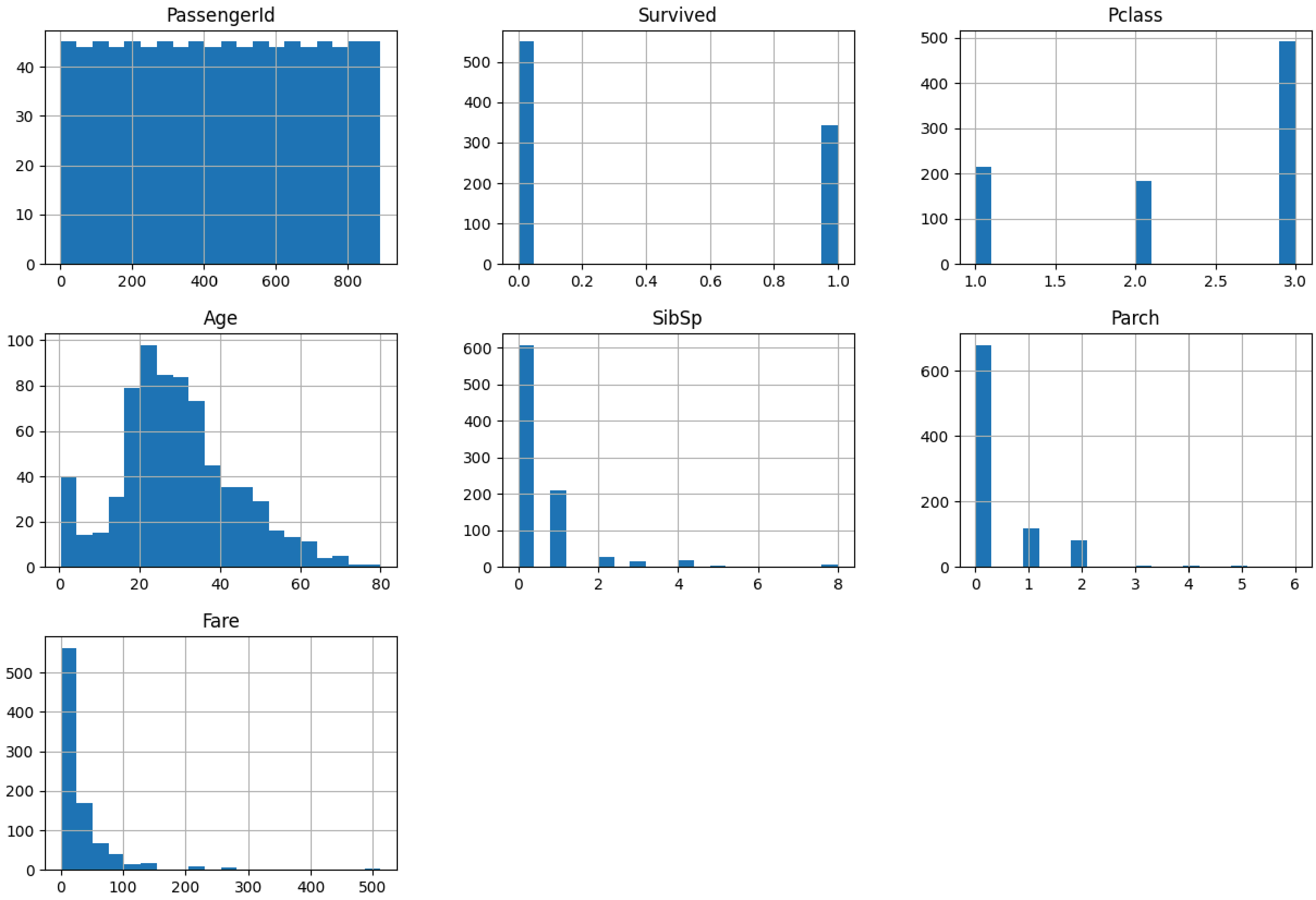
In [8]:
```
df.pd.read_csv("train.csv")
```

In [13]:
```
df.head()
df.info()
df.describe()
df.isnull().sum()
df['Survived'].value_counts()
```

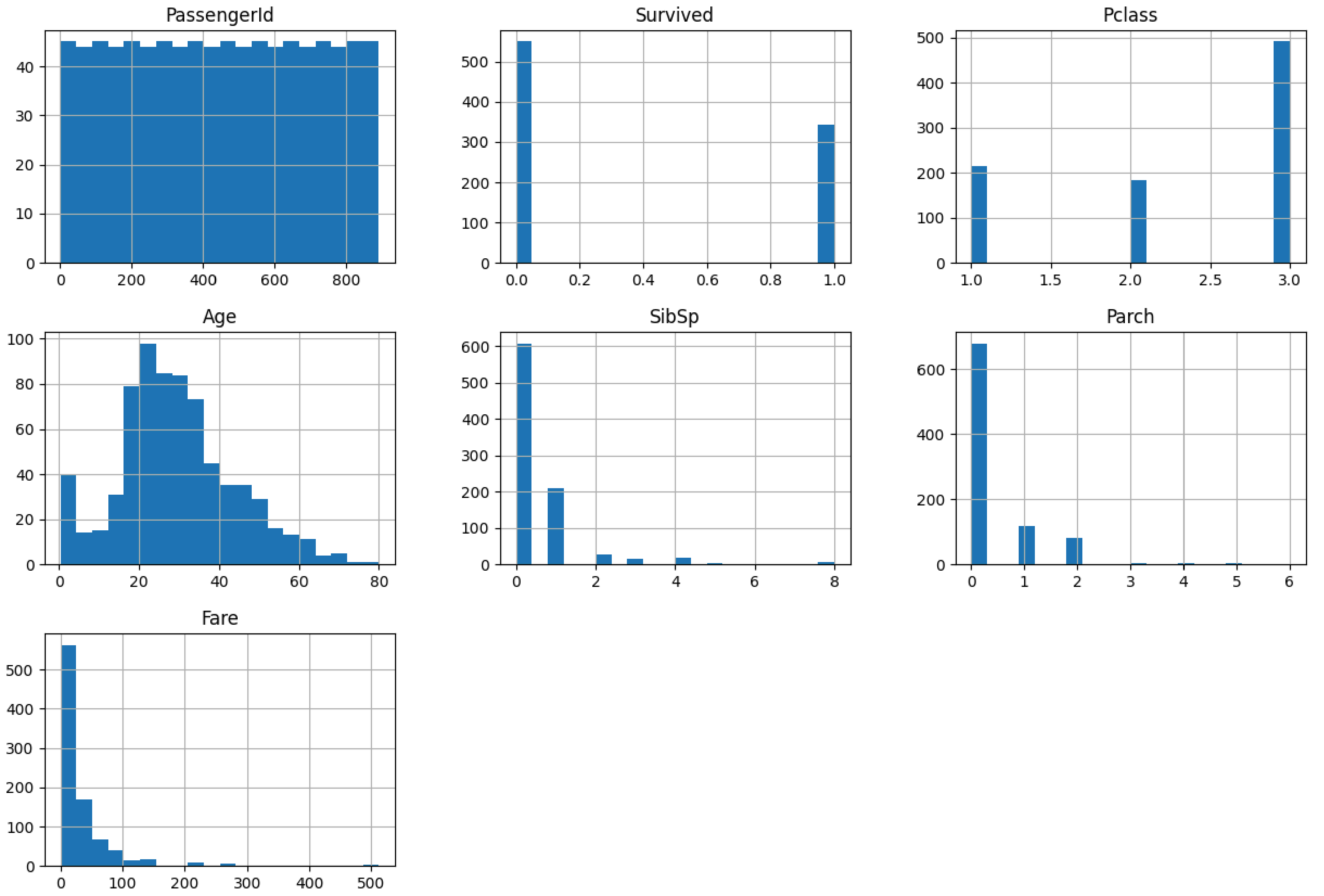
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    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
```

Out[13]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

In [12]:
```
df.hist(bins=20, figsize=(15,10))
plt.show()
```

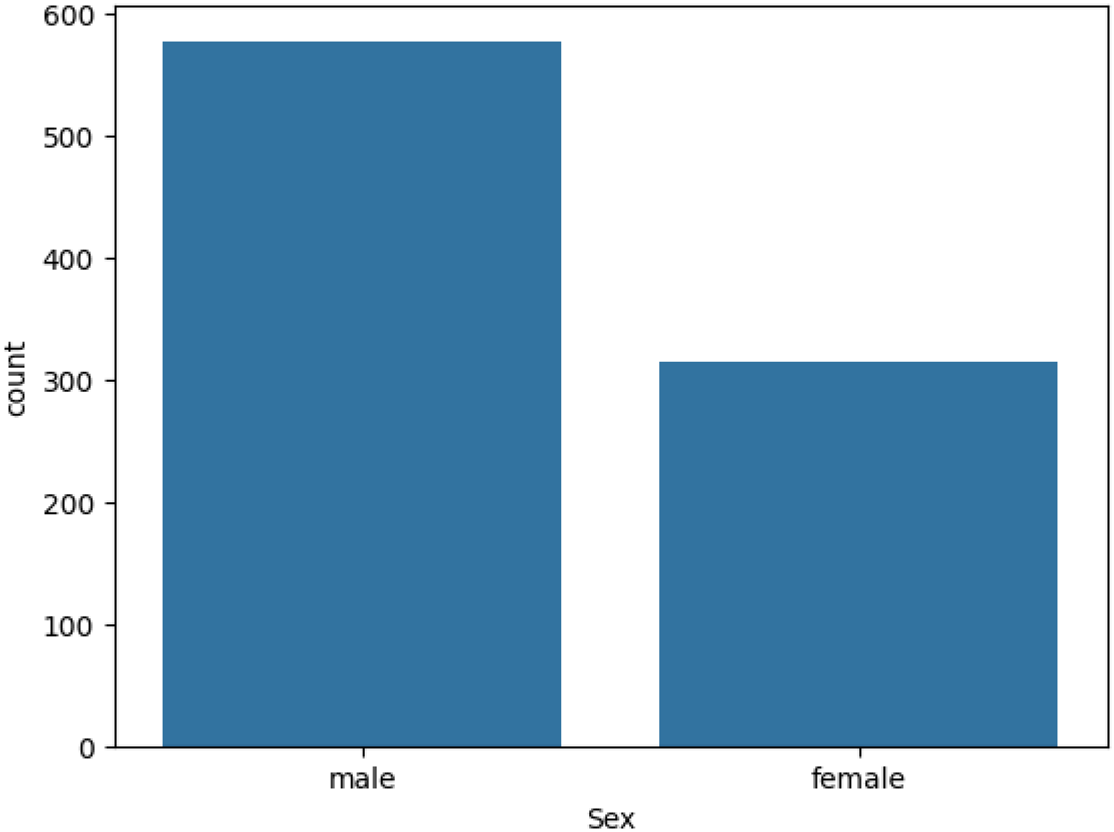


In [14]:
```
df.hist(bins=20, figsize=(15, 10))
plt.show()
```



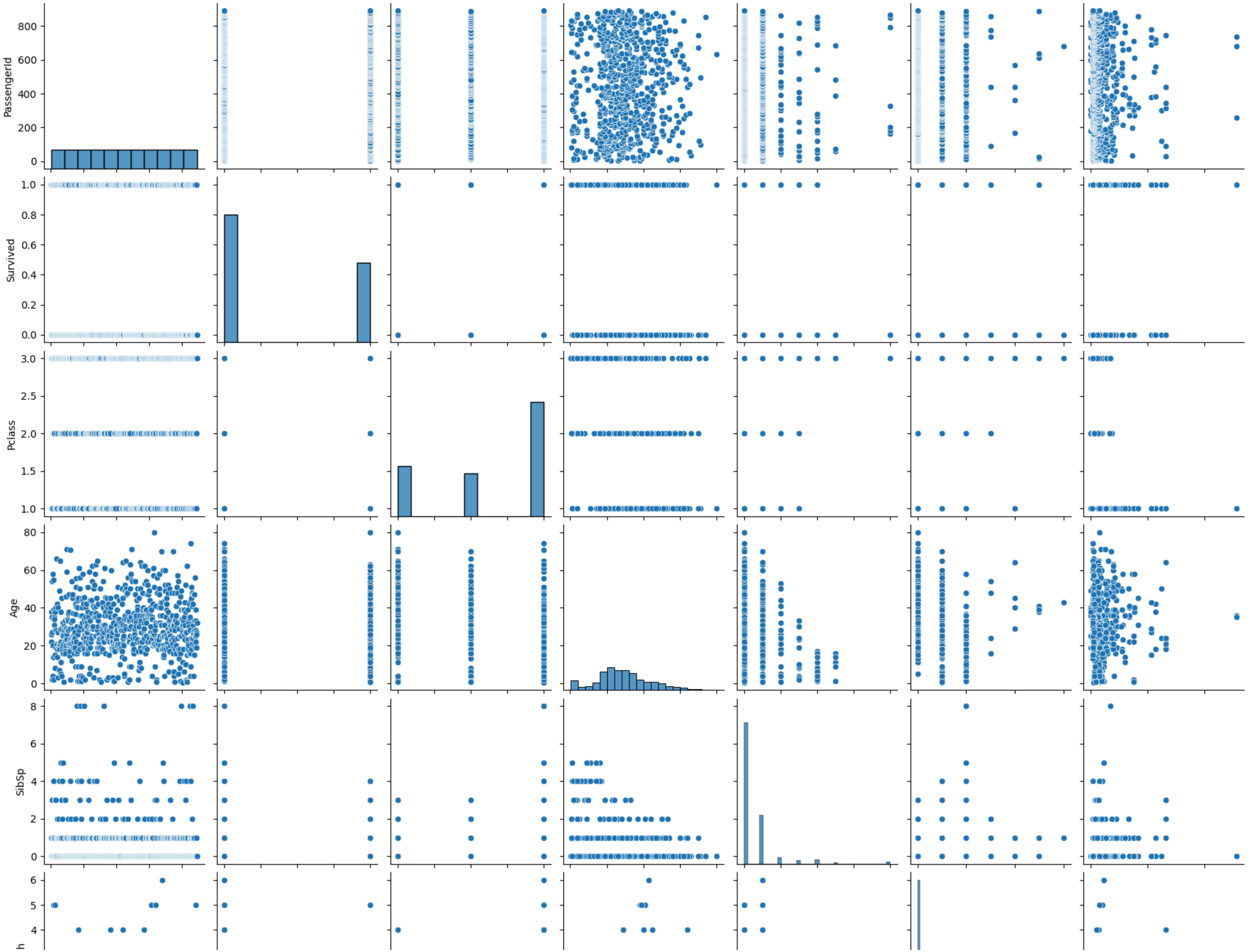
In [16]:
```
sns.countplot(x='Sex',data=df)
```

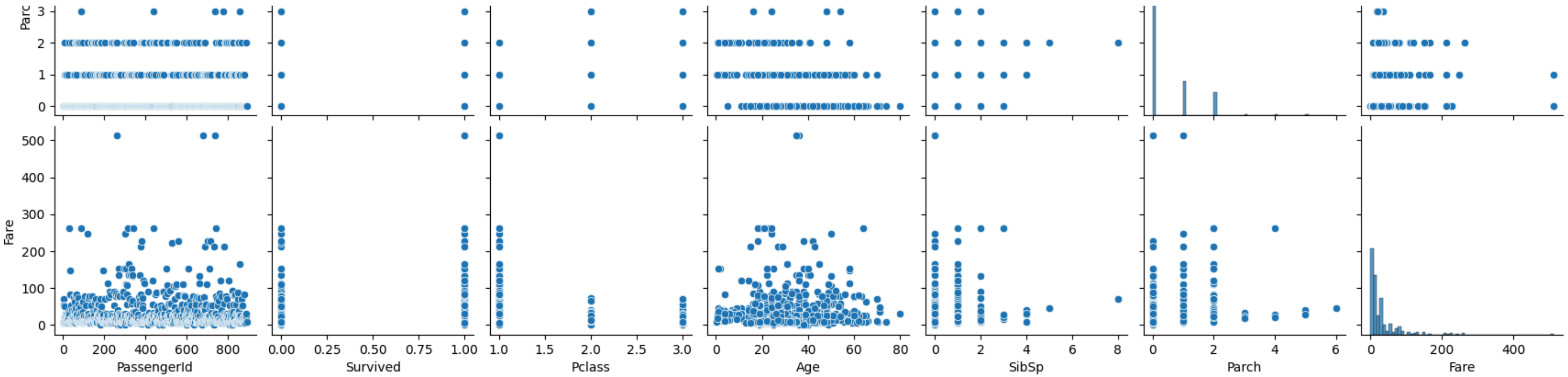
Out[16]: <Axes: xlabel='Sex', ylabel='count'>



In [19]:
```
sns.pairplot(df.select_dtypes(include=['int64','float64']))
```

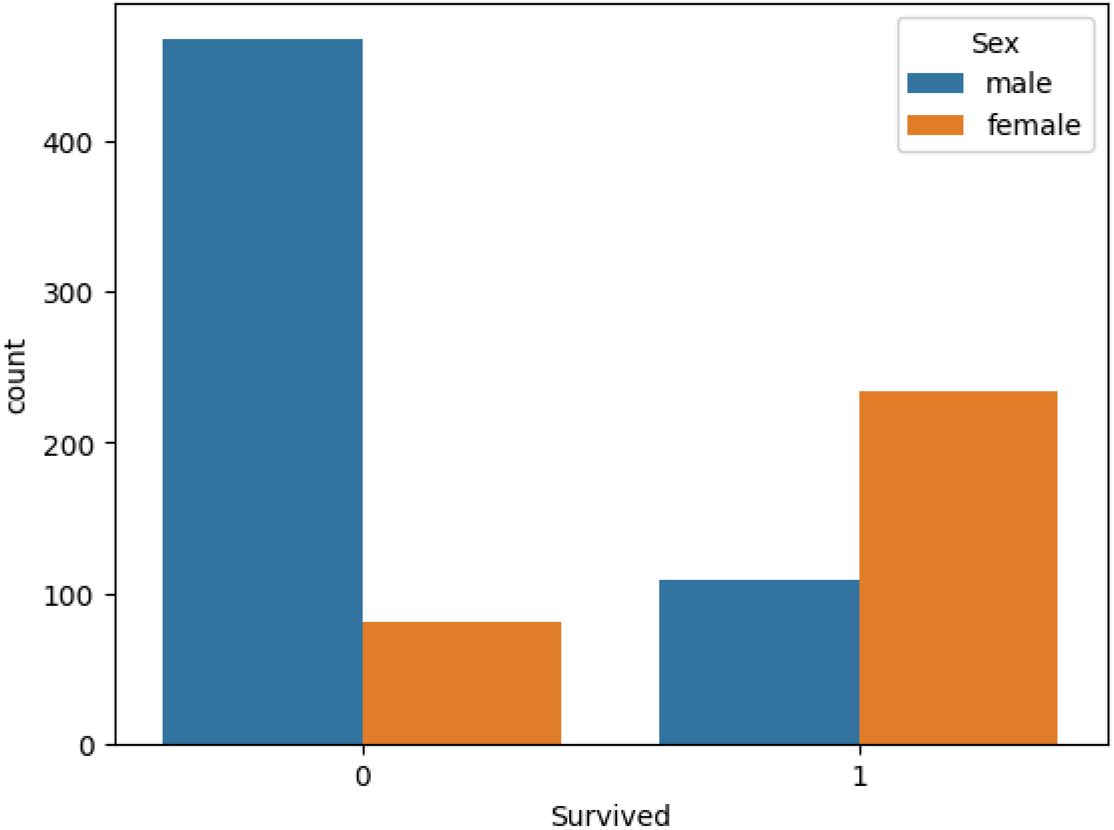
Out[19]: <seaborn.axisgrid.PairGrid at 0x122e65010>


PassengerId
800
600
400
200
0
Survived
1.0
0.8
0.6
0.4
0.2
0.0
Pclass
3.0
2.5
2.0
1.5
1.0
Age
80
60
40
20
0
SibSp
8
6
4
2
0
6
5
4



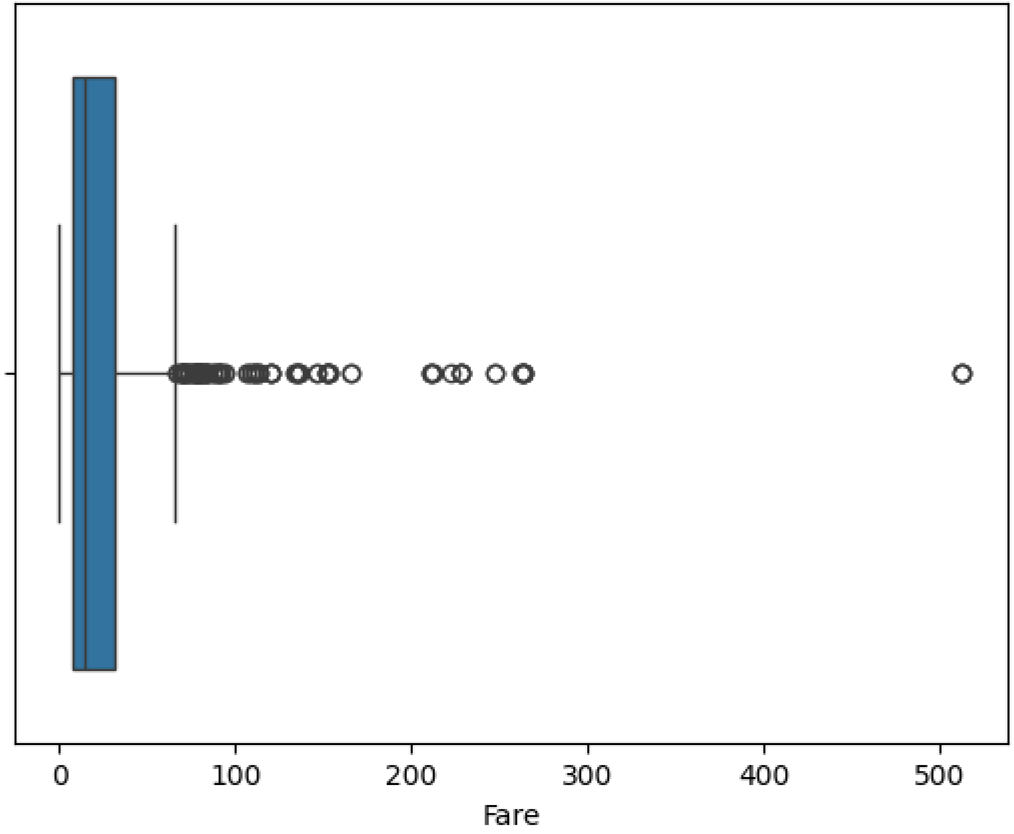
```
In [20]: sns.countplot(x='Survived', hue='Sex', data=df)
```

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



In [25]:
```
sns.boxplot(x=df['Fare'])
```

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



In [ ]:
```
## Summary of Findings

- The Titanic dataset contains 891 rows and 12 columns, including both numerical and categorical features.
- The `Age` column has around 20% missing values, while the `Cabin` column has about 77% missing values.
- Female passengers had a significantly higher survival rate (~74%) compared to male passengers (~19%).
- Passengers in 1st class had the highest survival rate, followed by 2nd class, and then 3rd class.
- Younger passengers, especially children, had better chances of survival compared to older passengers.
- Fare is strongly related to passenger class — higher fares are typically from 1st-class tickets.
- Outliers are present in the `Fare` column, indicating that a few passengers paid significantly higher ticket pric
- The heatmap shows that `Fare` and `Pclass` are negatively correlated, and `Sex` has a strong relationship with `S
```

In [ ]:

In [ ]:

```
```