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

```python
In [2]:  df = sns.load_dataset('titanic')
```

```python
In [3]:  # View general structure and stats
         print(df.info())
         print(df.describe())
         print(df['survived'].value_counts())
         print(df.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
None
         survived      pclass         age       sibsp       parch        fare
count  891.000000  891.000000  714.000000  891.000000  891.000000  891.000000
mean     0.383838    2.308642   29.699118    0.523008    0.381594   32.204208
std      0.486592    0.836071   14.526497    1.102743    0.806057   49.693429
min      0.000000    1.000000    0.420000    0.000000    0.000000    0.000000
25%      0.000000    2.000000   20.125000    0.000000    0.000000    7.910400
50%      0.000000    3.000000   28.000000    0.000000    0.000000   14.454200
75%      1.000000    3.000000   38.000000    1.000000    0.000000   31.000000
max      1.000000    3.000000   80.000000    8.000000    6.000000  512.329200
survived
0    549
1    342
Name: count, dtype: int64
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
```
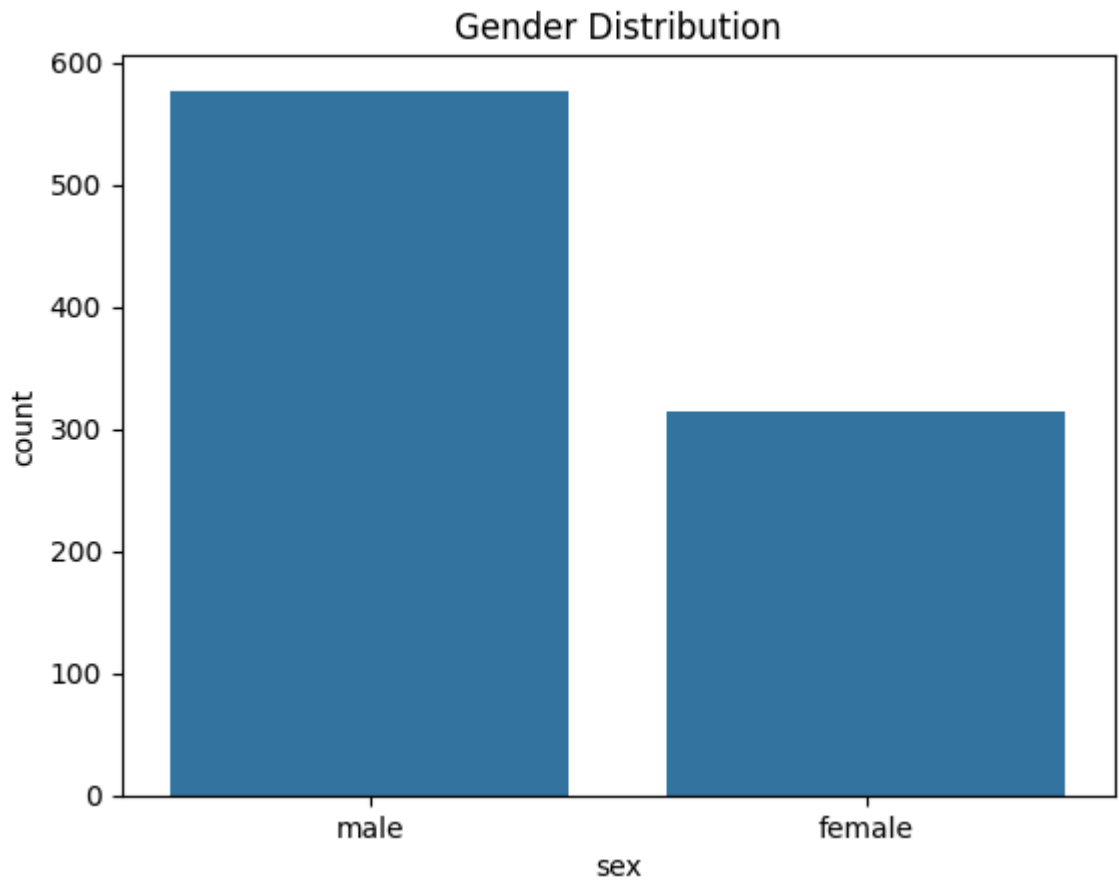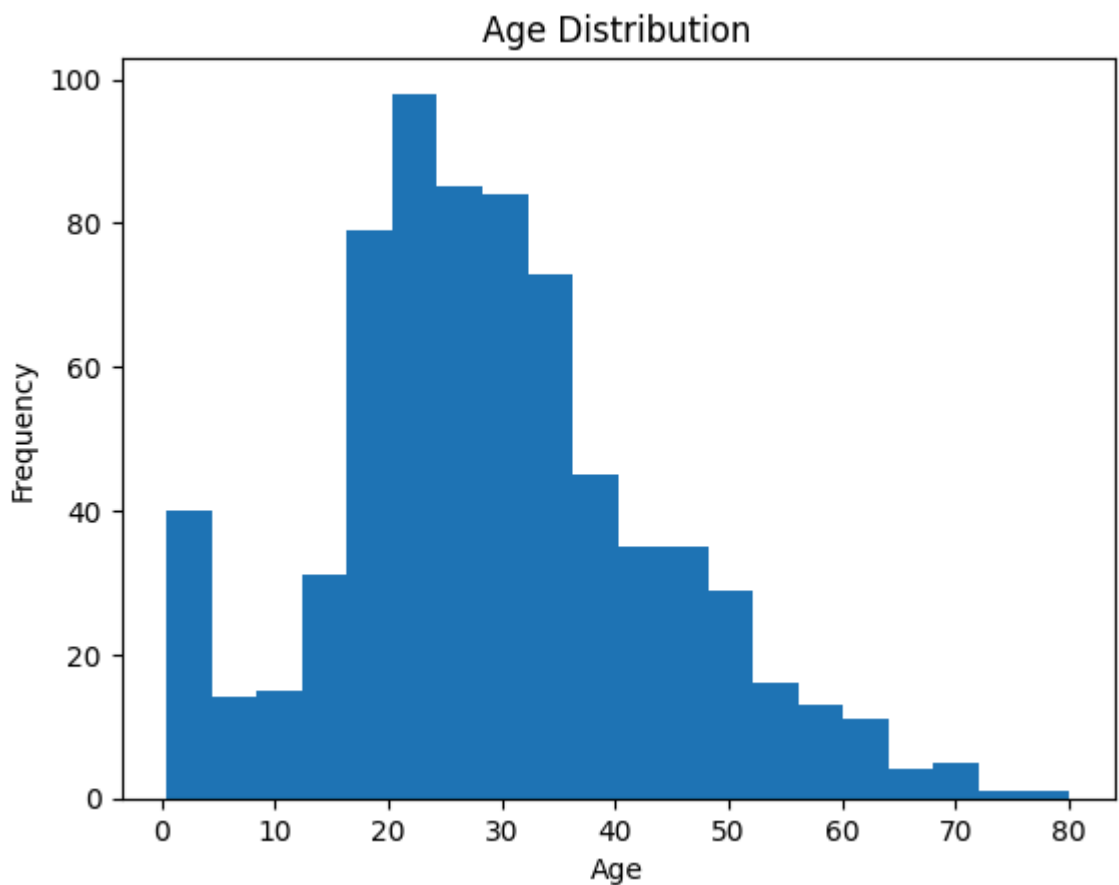
In [4]:
```python
# Gender distribution
sns.countplot(x='sex', data=df)
plt.title('Gender Distribution')
plt.show()
```
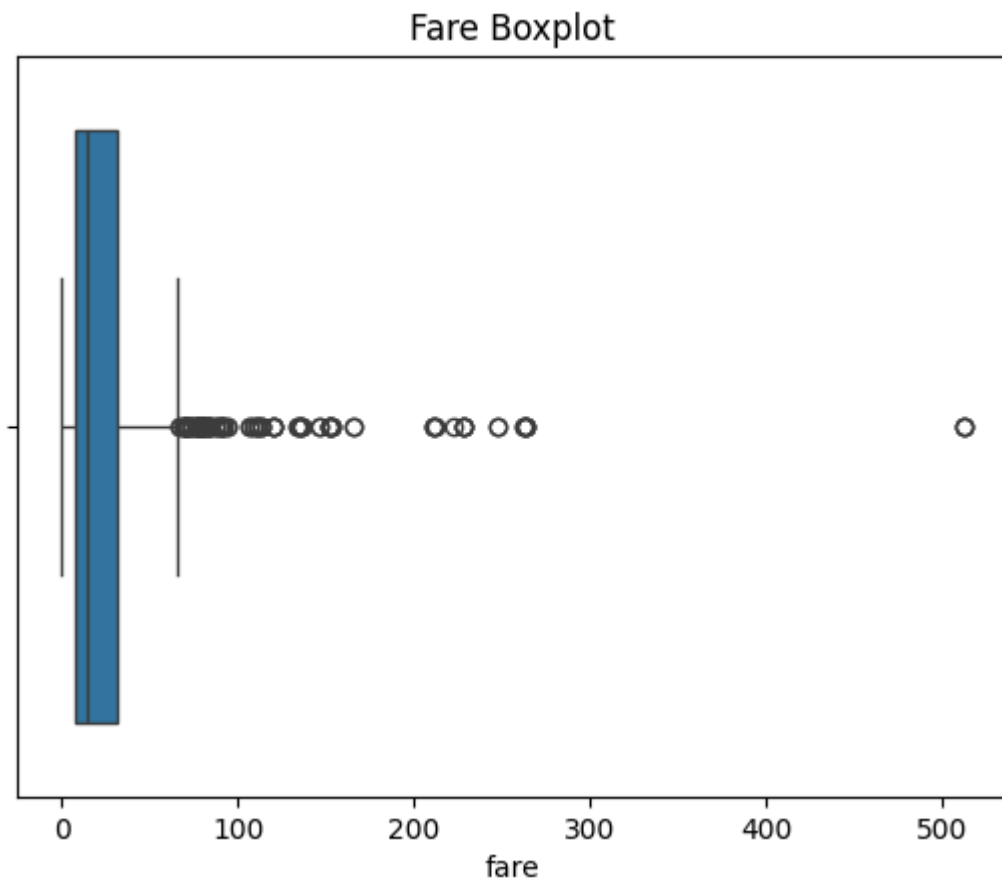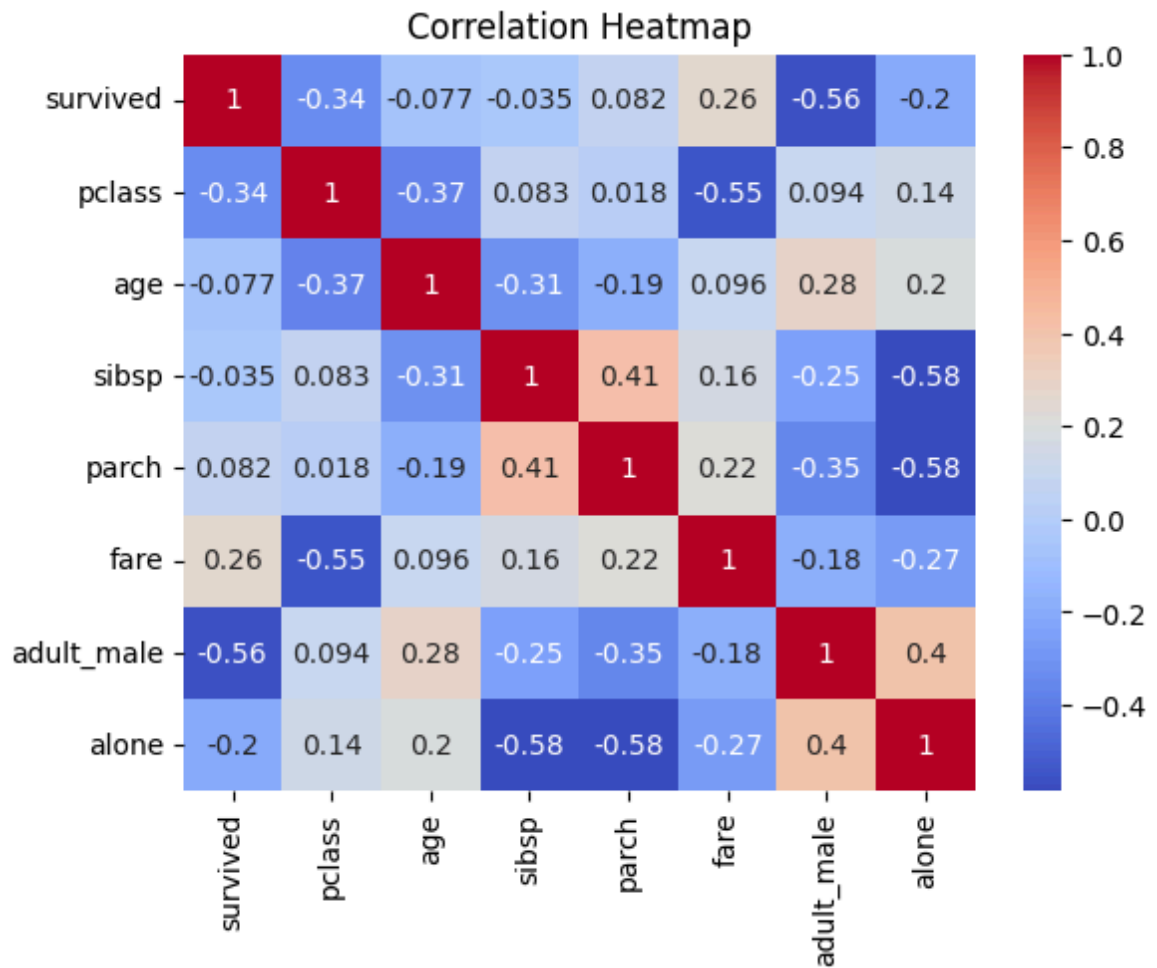
Gender Distribution


```
In [5]: # Age distribution
        df['age'].plot(kind='hist', bins=20, title='Age Distribution')
        plt.xlabel('Age')
        plt.show()
```
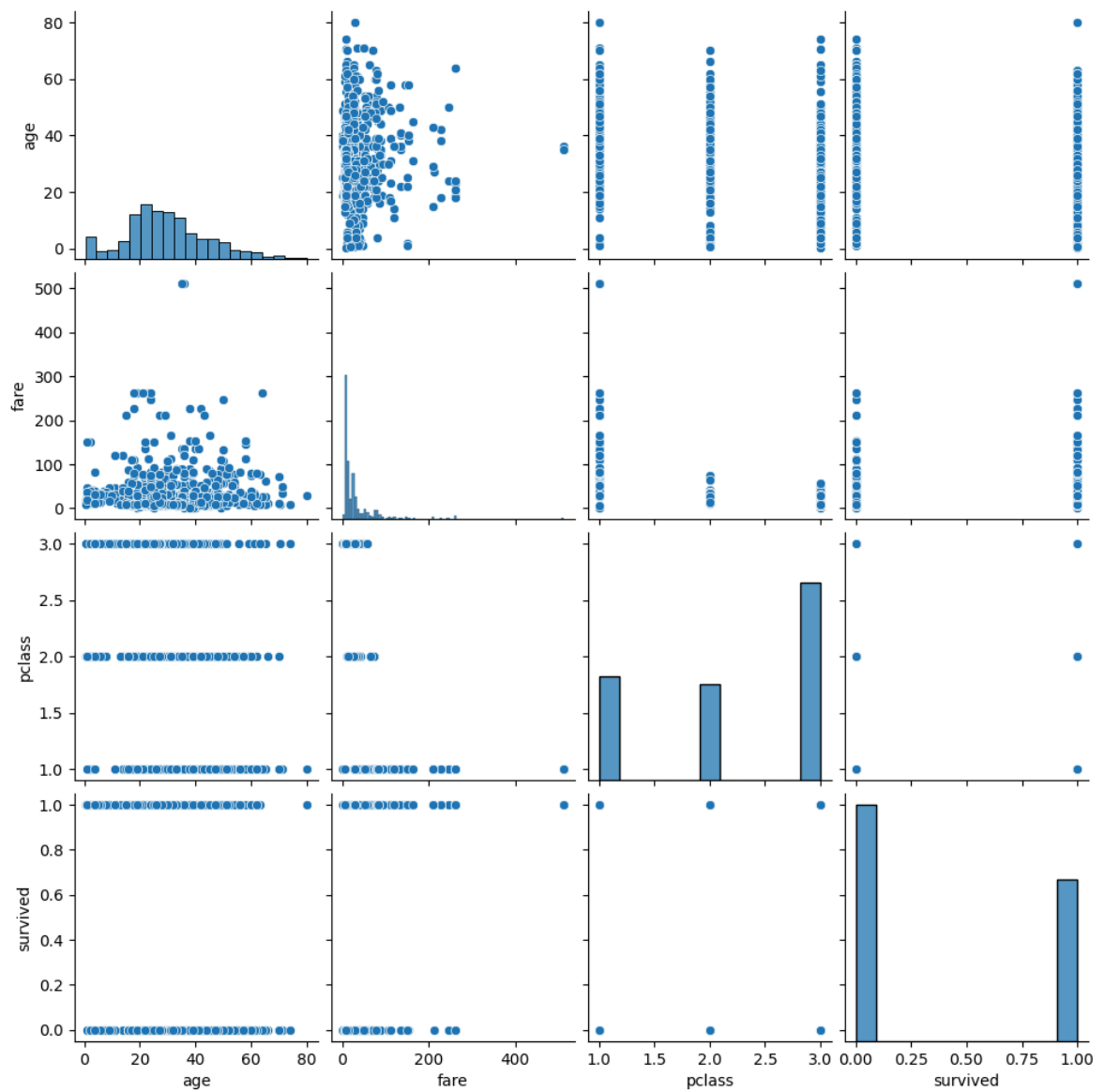
Age Distribution

In [6]:
```python
# Fare boxplot
sns.boxplot(x='fare', data=df)
plt.title('Fare Boxplot')
plt.show()
```

Fare Boxplot

In [7]:
```python
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

Correlation Heatmap

|	survived	pclass	age	sibsp	parch	fare	adult_male	alone
survived	1	-0.34	-0.077	-0.035	0.082	0.26	-0.56	-0.2
pclass	-0.34	1	-0.37	0.083	0.018	-0.55	0.094	0.14
age	-0.077	-0.37	1	-0.31	-0.19	0.096	0.28	0.2
sibsp	-0.035	0.083	-0.31	1	0.41	0.16	-0.25	-0.58
parch	0.082	0.018	-0.19	0.41	1	0.22	-0.35	-0.58
fare	0.26	-0.55	0.096	0.16	0.22	1	-0.18	-0.27
adult_male	-0.56	0.094	0.28	-0.25	-0.35	-0.18	1	0.4
alone	-0.2	0.14	0.2	-0.58	-0.58	-0.27	0.4	1

In [8]:
```python
# Pairplot
sns.pairplot(df[['age', 'fare', 'pclass', 'survived']].dropna())
plt.show()
```

```python
# Survival by gender
sns.barplot(x='sex', y='survived', data=df)
plt.title('Survival Rate by Gender')
plt.show()
```

Survival Rate by Gender
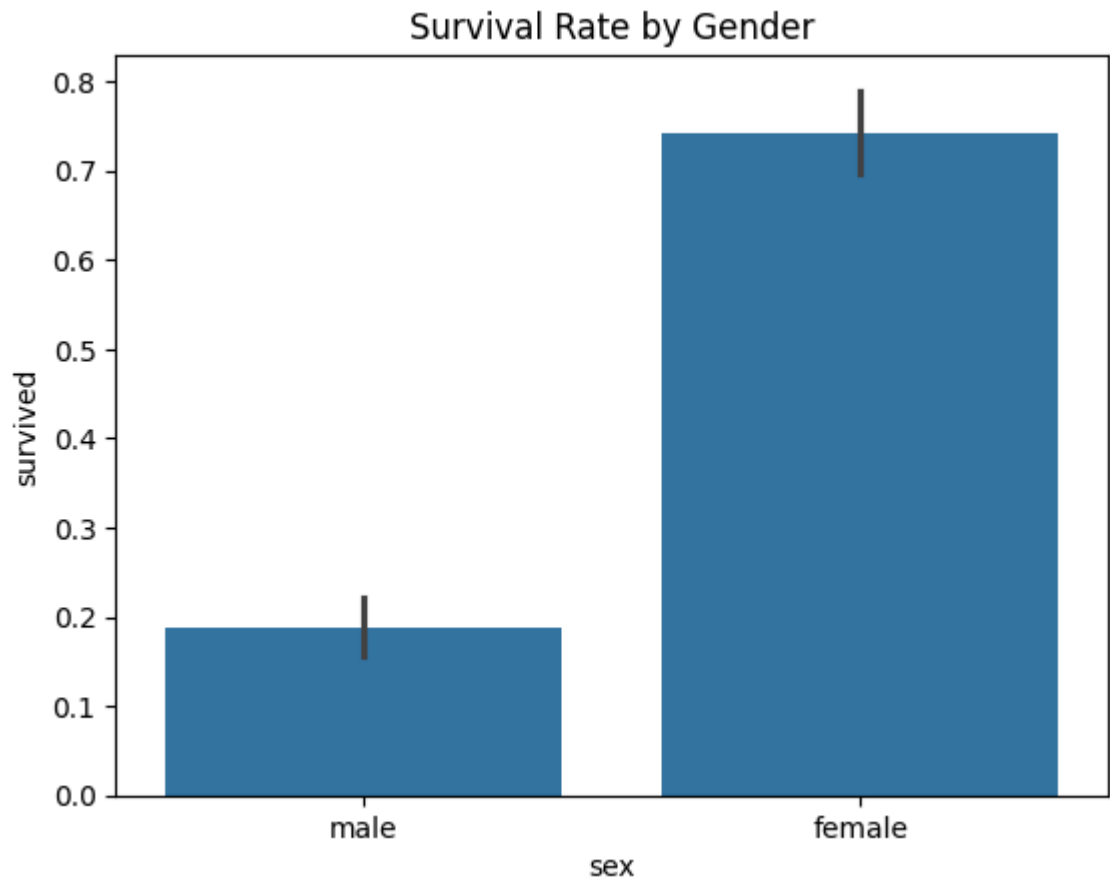
In [10]:
```python
# Gender vs Survival across Classes
sns.catplot(x='sex', hue='survived', col='pclass', data=df, kind='count')
plt.show()
```
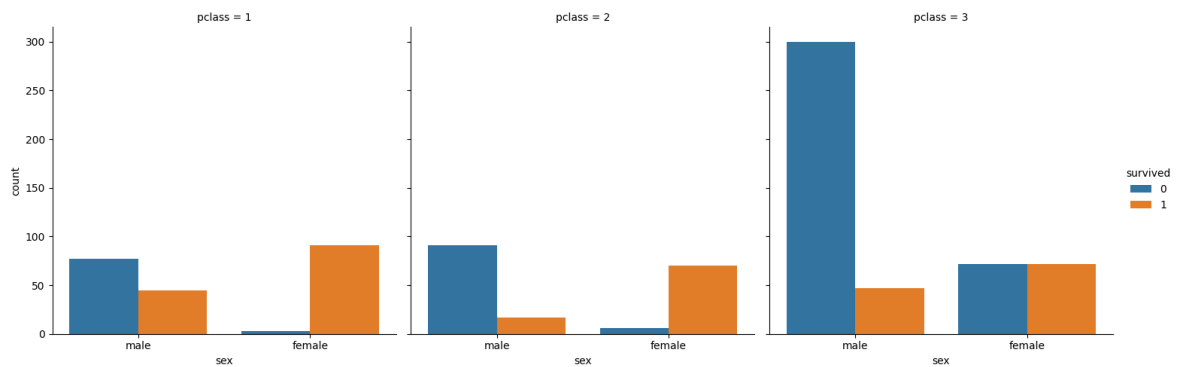

In [11]:
```python
# Skewness check and transformation
print("Fare Skewness:", df['fare'].skew())
df['fare_log'] = np.log1p(df['fare'])
print("Fare_log Skewness:", df['fare_log'].skew())
```

```
Fare Skewness: 4.787316519674893
Fare_log Skewness: 0.3949280095189306
```