```
In [7]:   # Import libraries
          from google.colab import files
          import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns    # <-- this is the missing part

          # Upload train.csv
          uploaded = files.upload()

          # Load dataset
          df = pd.read_csv("train.csv")
          df.head()
```

Upload widget is only available when the cell has been executed
in the current browser session. Please rerun this cell to enable.
Saving train.csv to train (1).csv

Out[7]:

	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
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
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450

```
In [8]:   # Dataset overview
          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 [9]: `# Summary of numerical features`
`df.describe()`

Out[9]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000

In [10]: `# Check missing values`
`df.isnull().sum()`

Out[10]:

	0
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 [11]:
```python
# Count unique values in each column
df.nunique()
```

Out[11]:

	0
PassengerId	891
Survived	2
Pclass	3
Name	891
Sex	2
Age	88
SibSp	7
Parch	7
Ticket	681
Fare	248
Cabin	147
Embarked	3

dtype: int64

In [12]:
```python
sns.histplot(df['Age'].dropna(), kde=True)
plt.title("Age Distribution")
plt.show()
```


Age Distribution
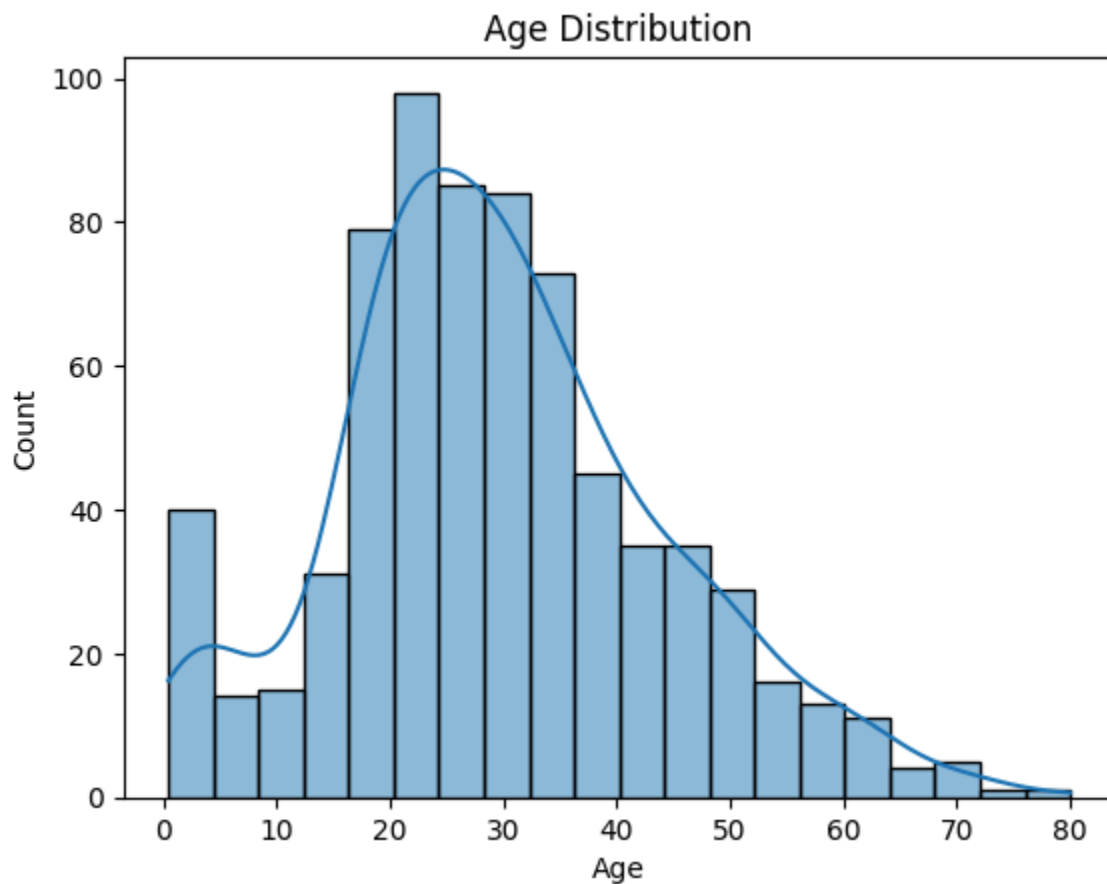
In [13]:
```python
sns.countplot(x='Survived', data=df)
plt.title("Survival Count (0 = No, 1 = Yes)")
plt.show()
```

Survival Count (0 = No, 1 = Yes)
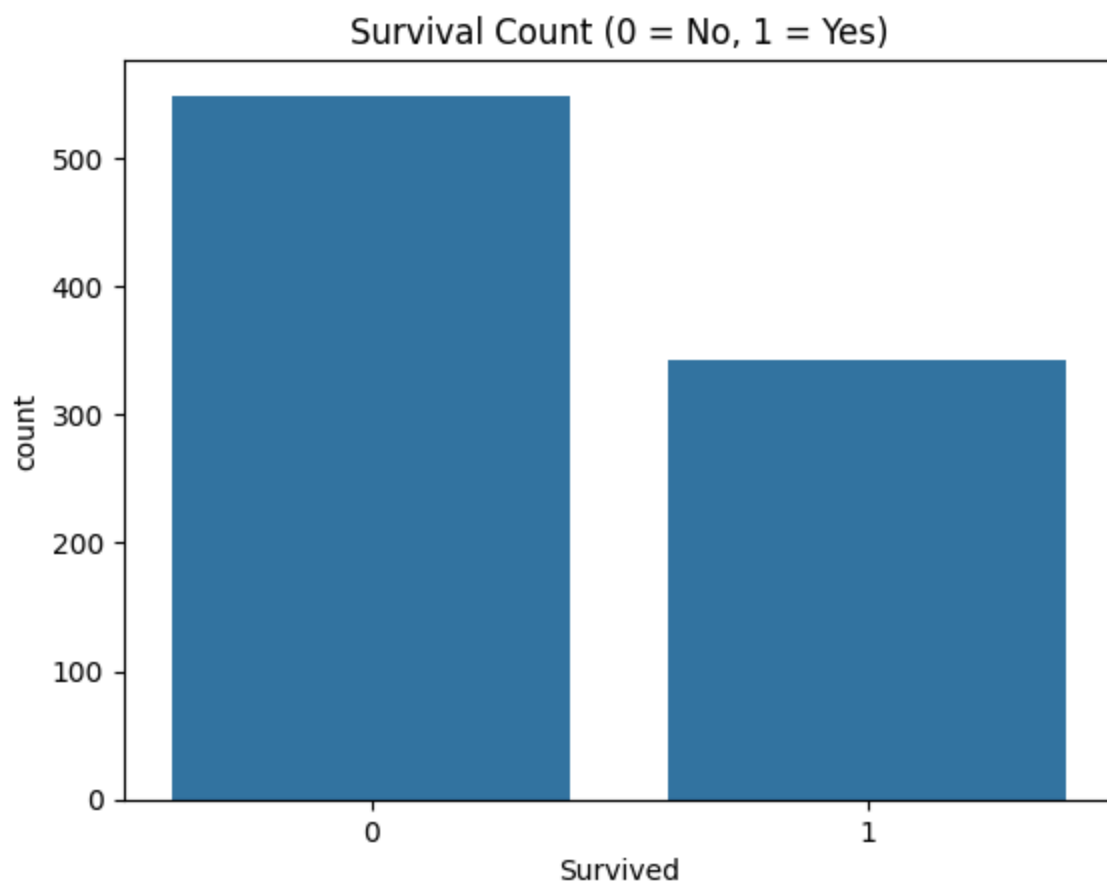
```
In [14]:  sns.countplot(x='Sex', hue='Survived', data=df)
          plt.title("Survival by Gender")
          plt.show()
```

Survival by Gender
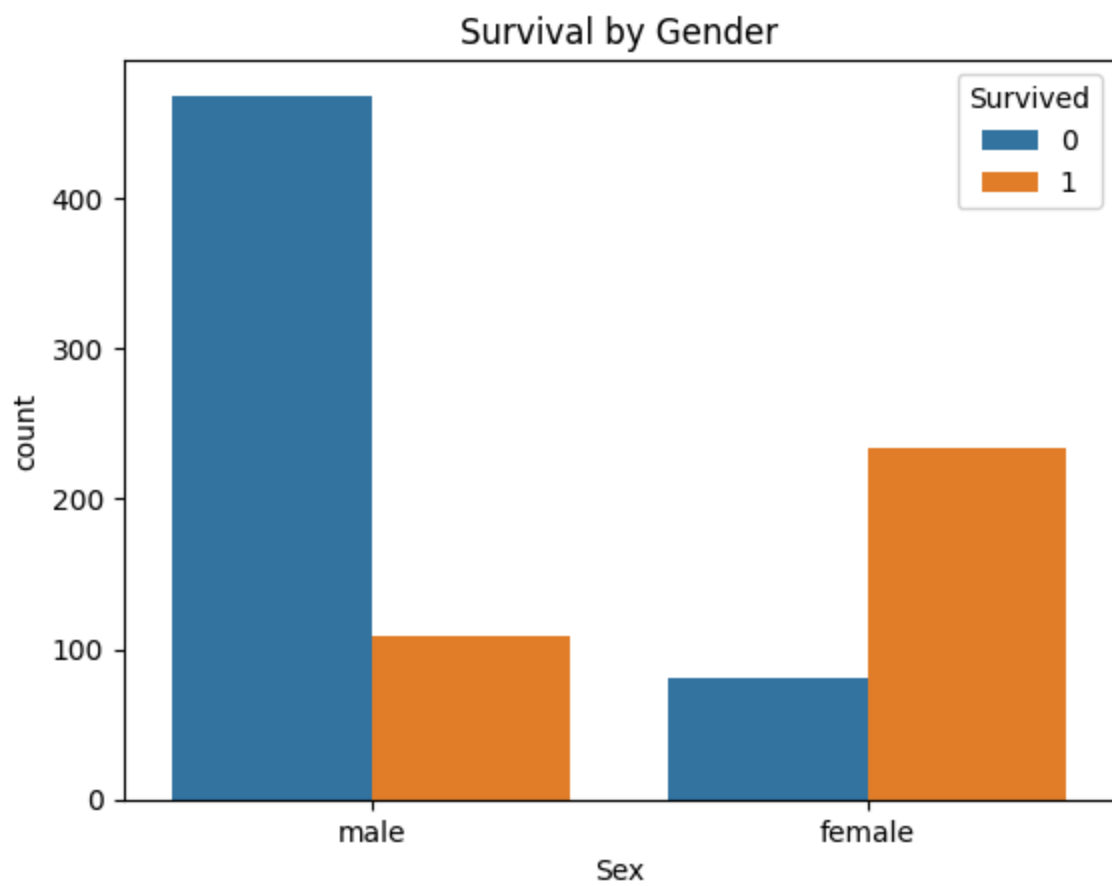
```
In [15]: sns.countplot(x='Pclass', hue='Survived', data=df)
         plt.title("Survival by Passenger Class")
         plt.show()
```

Survival by Passenger Class
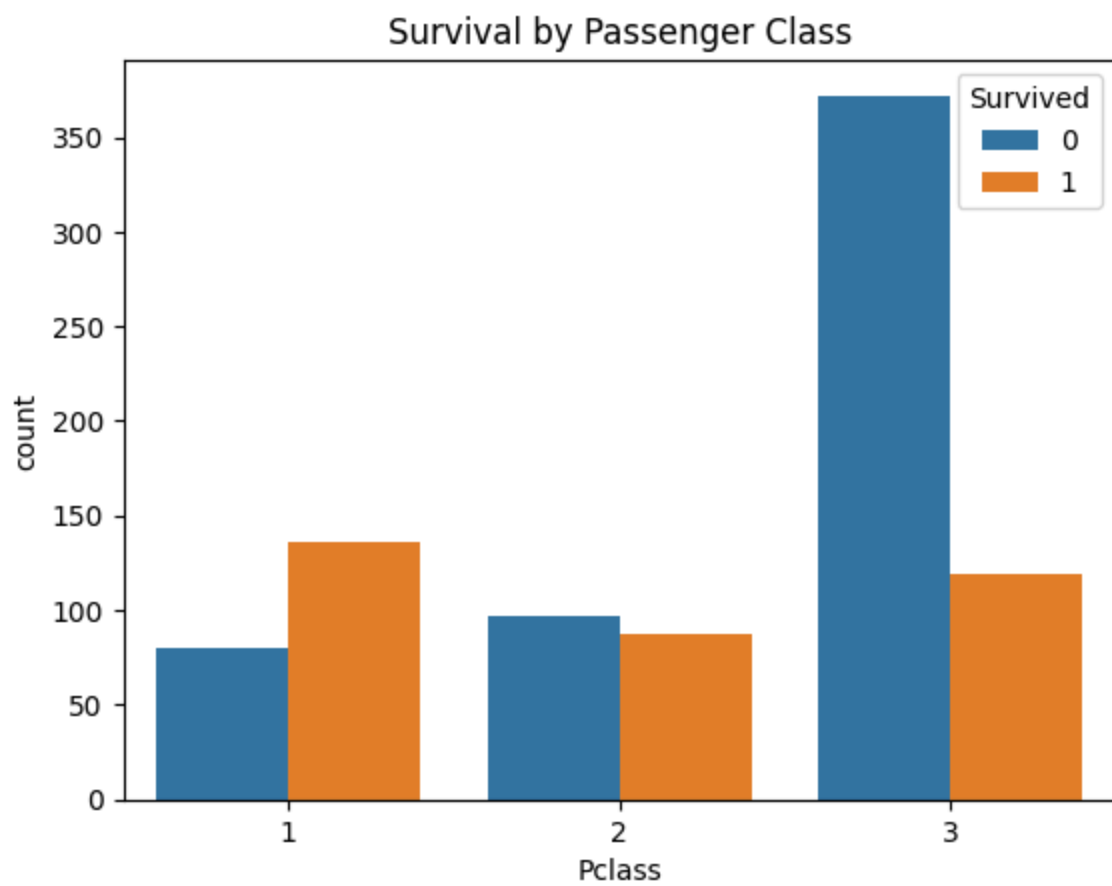
In [16]:
```python
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
plt.title("Age vs Fare (colored by Survival)")
plt.show()
```
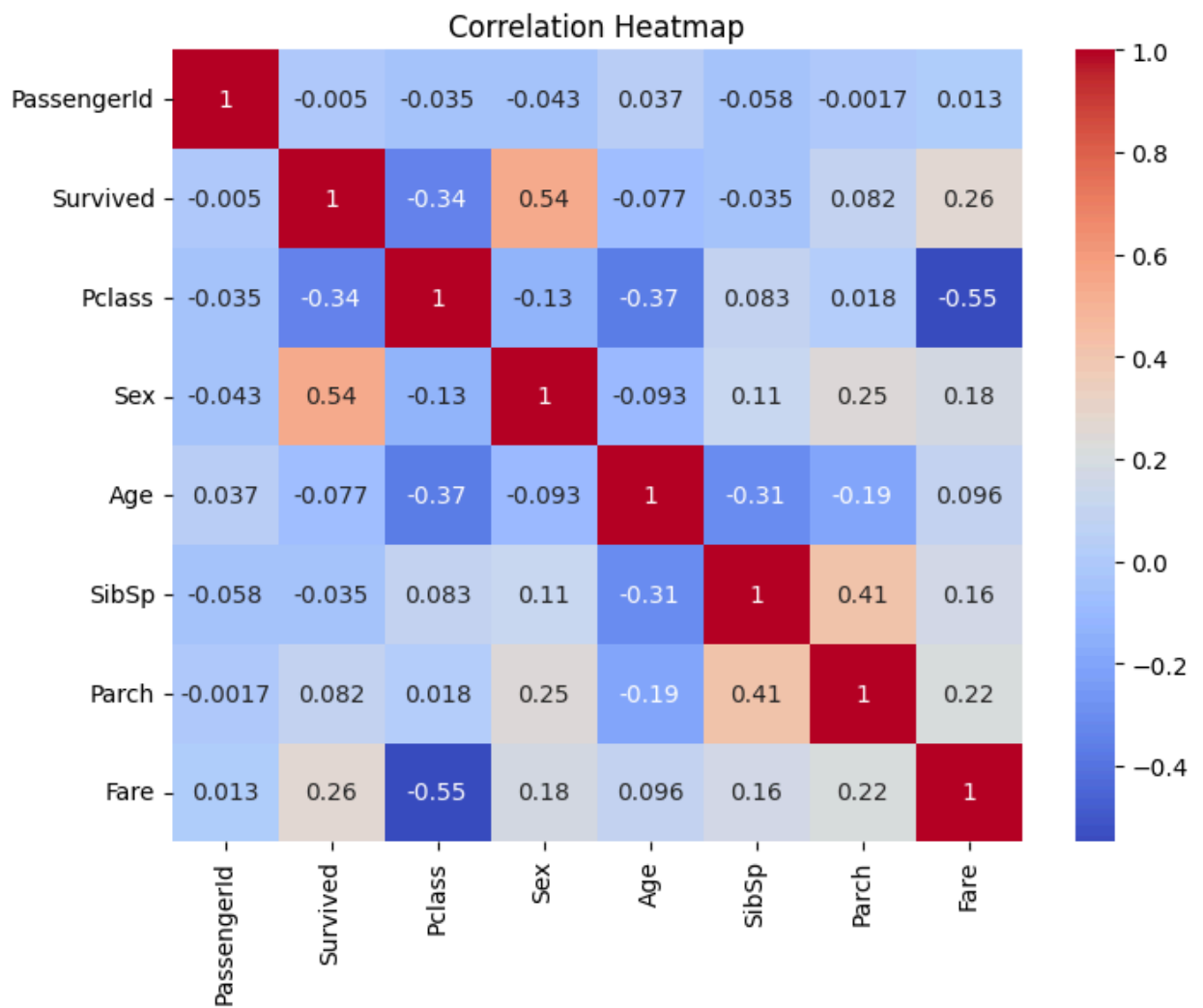
Age vs Fare (colored by Survival)
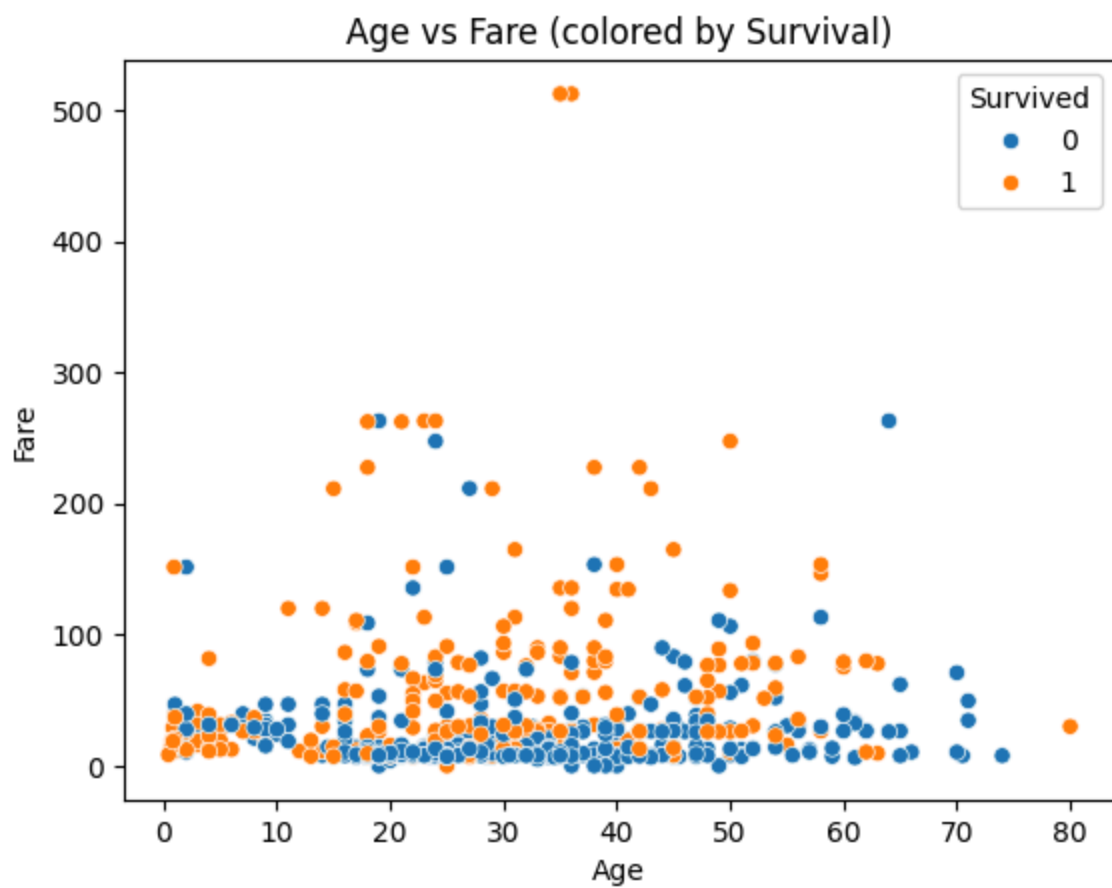
```
In [17]:  sns.boxplot(x='Survived', y='Age', data=df)
          plt.title("Age vs Survival")
          plt.show()
```

Age vs Survival
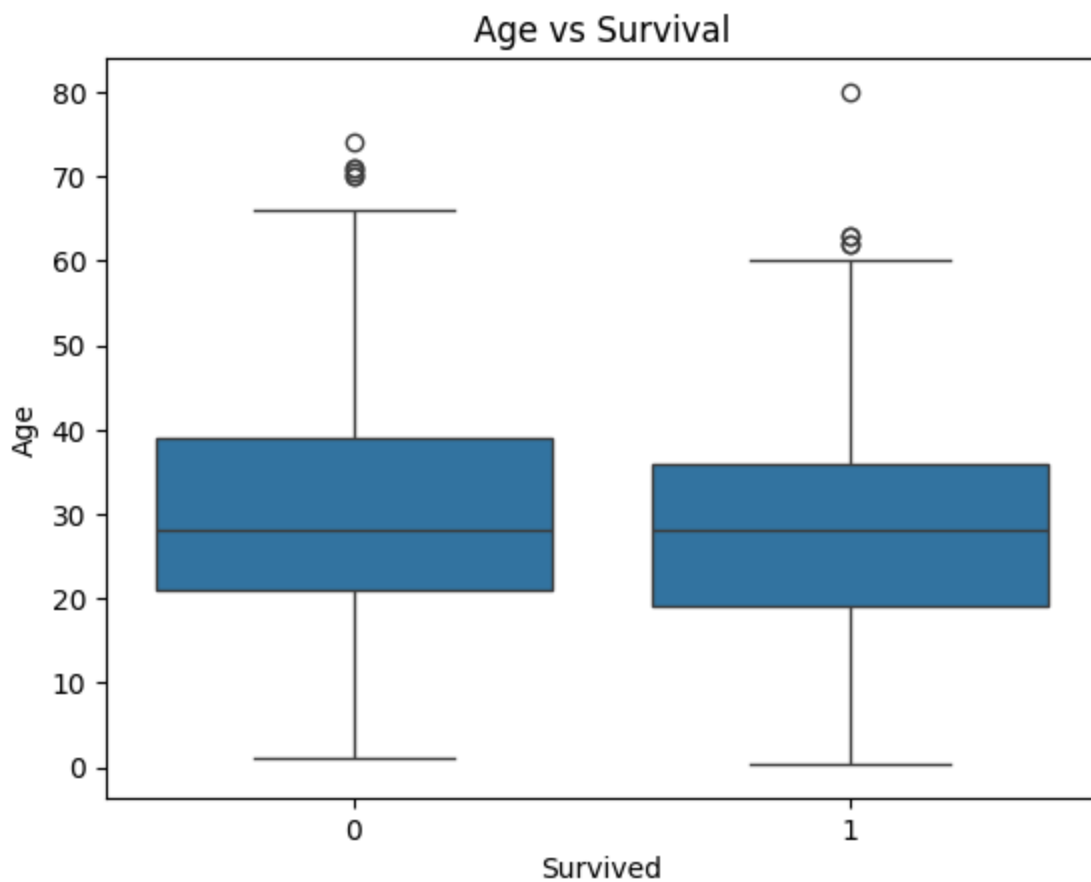
In [19]:
```python
# Encode Sex column for correlation
df_encoded = df.copy()
df_encoded['Sex'] = df_encoded['Sex'].map({'male': 0, 'female': 1})

# Select only numeric columns
numeric_df = df_encoded.select_dtypes(include=['int64', 'float64'])

# Correlation heatmap
plt.figure(figsize=(8,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```
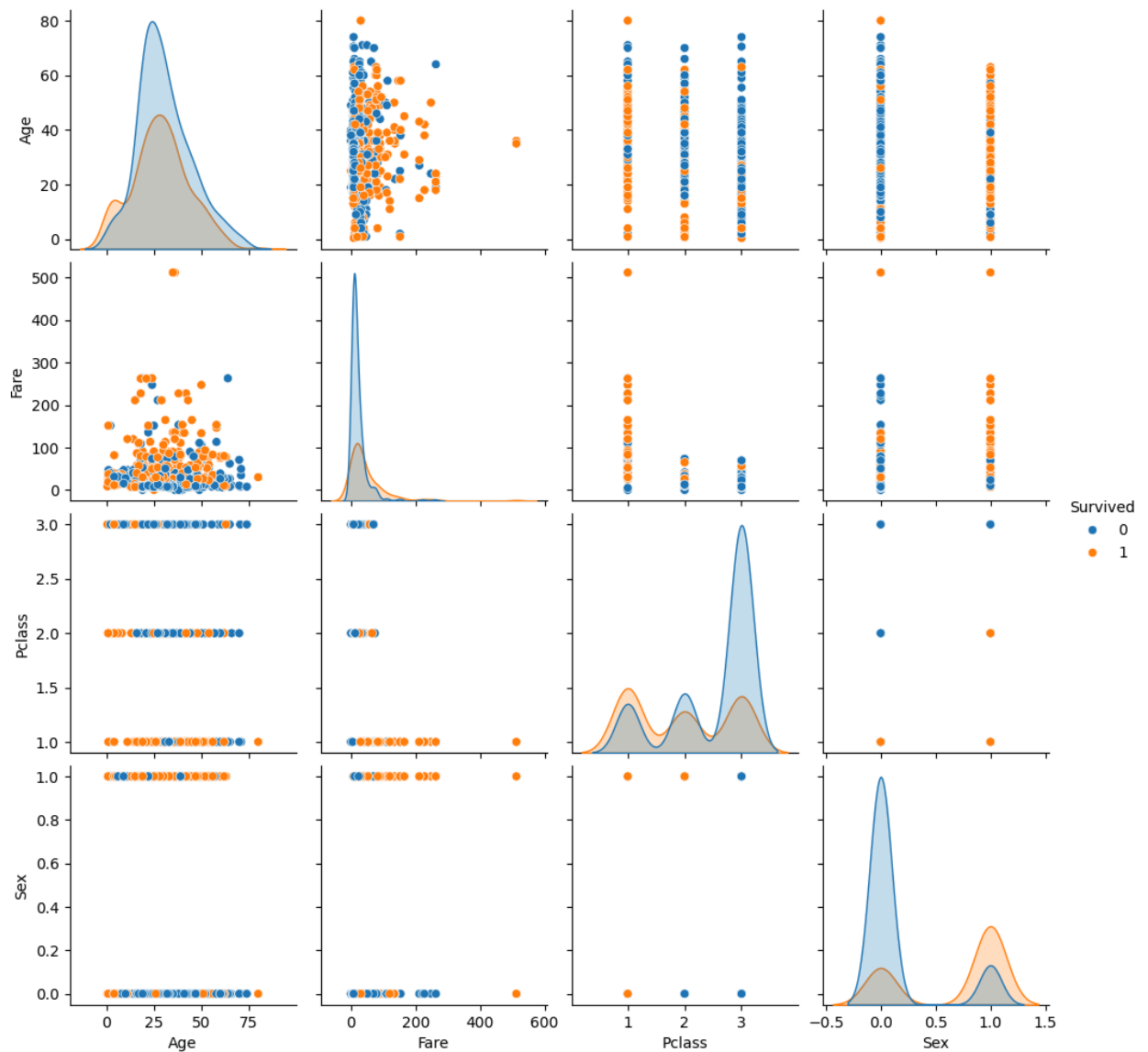
Correlation Heatmap

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
PassengerId	1	-0.005	-0.035	-0.043	0.037	-0.058	-0.0017	0.013
Survived	-0.005	1	-0.34	0.54	-0.077	-0.035	0.082	0.26
Pclass	-0.035	-0.34	1	-0.13	-0.37	0.083	0.018	-0.55
Sex	-0.043	0.54	-0.13	1	-0.093	0.11	0.25	0.18
Age	0.037	-0.077	-0.37	-0.093	1	-0.31	-0.19	0.096
SibSp	-0.058	-0.035	0.083	0.11	-0.31	1	0.41	0.16
Parch	-0.0017	0.082	0.018	0.25	-0.19	0.41	1	0.22
Fare	0.013	0.26	-0.55	0.18	0.096	0.16	0.22	1

In [20]:
```python
sns.pairplot(df_encoded[['Survived','Age','Fare','Pclass','Sex']], hue='Surviv
plt.show()
```

Summary of Insights

- **Gender**: Females had a much higher survival rate compared to males.
- **Class**: Passengers in 1st class had significantly better survival chances than those in 3rd class.
- **Age**: Younger passengers (especially children) had slightly better survival rates compared to older passengers.
- **Fare**: Higher ticket fares were positively correlated with survival (wealthier passengers survived more).
- **Overall**: Survival strongly depended on **social status (class)** and **gender**. These two were the most important factors in predicting survival.