In [4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [6]:
```
df = pd.read_csv("train.csv")
df.head()
```

Out[6]:

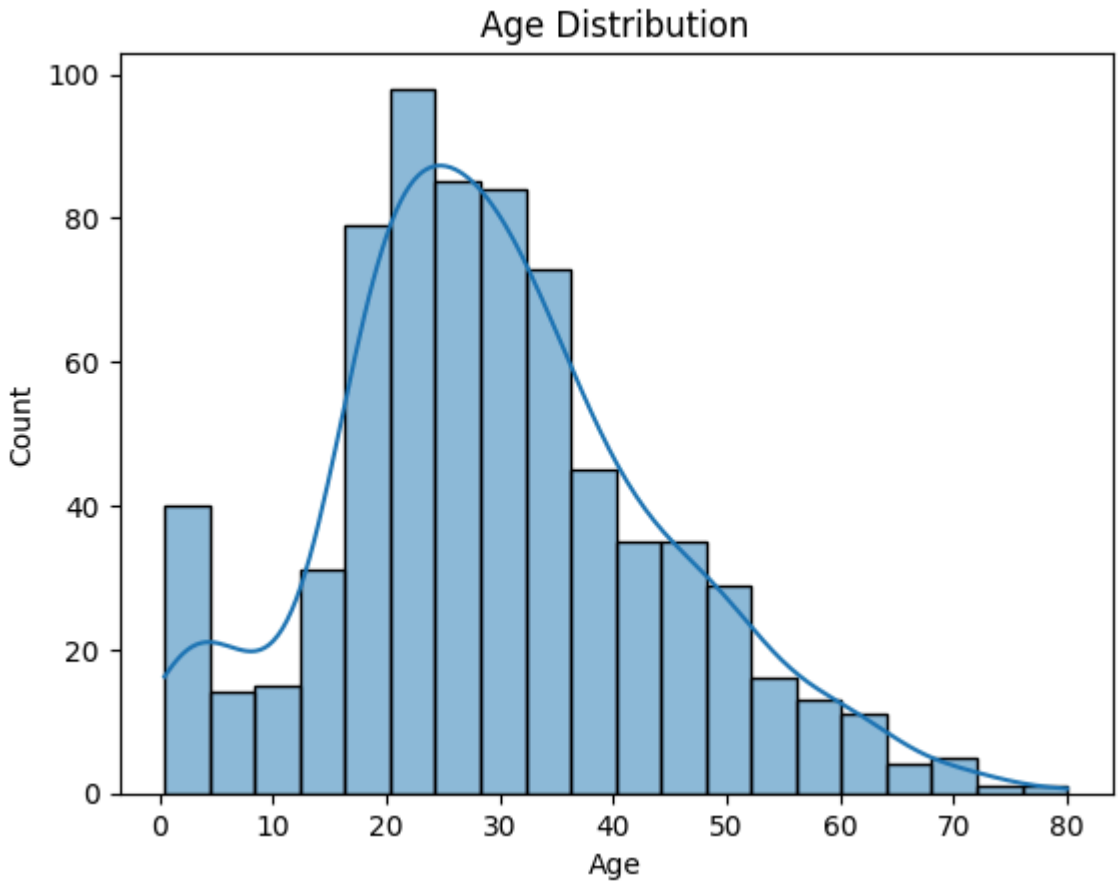
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques 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 [7]:
```
df.shape          # rows, columns
df.info()         # datatypes + null values
df.describe()     # summary statistics
df.isnull().sum() # missing values
df['Survived'].value_counts()
```

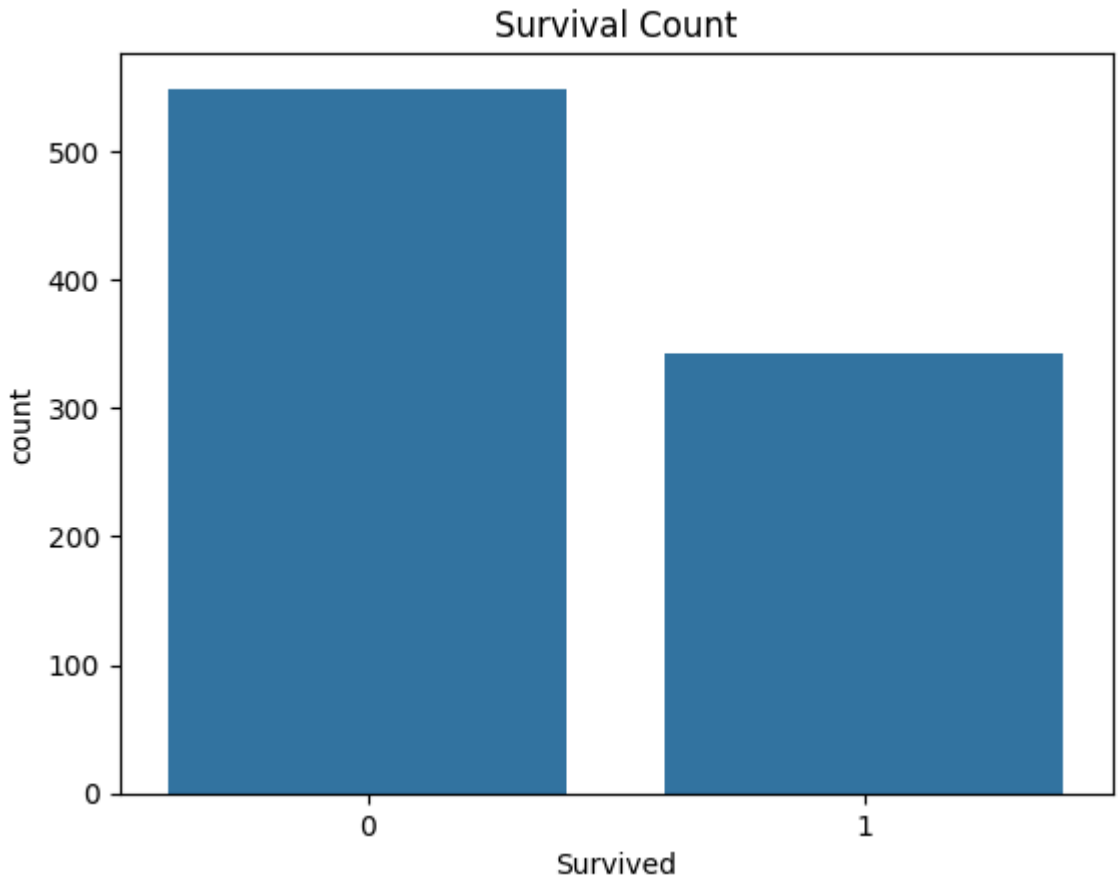
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    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[7]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

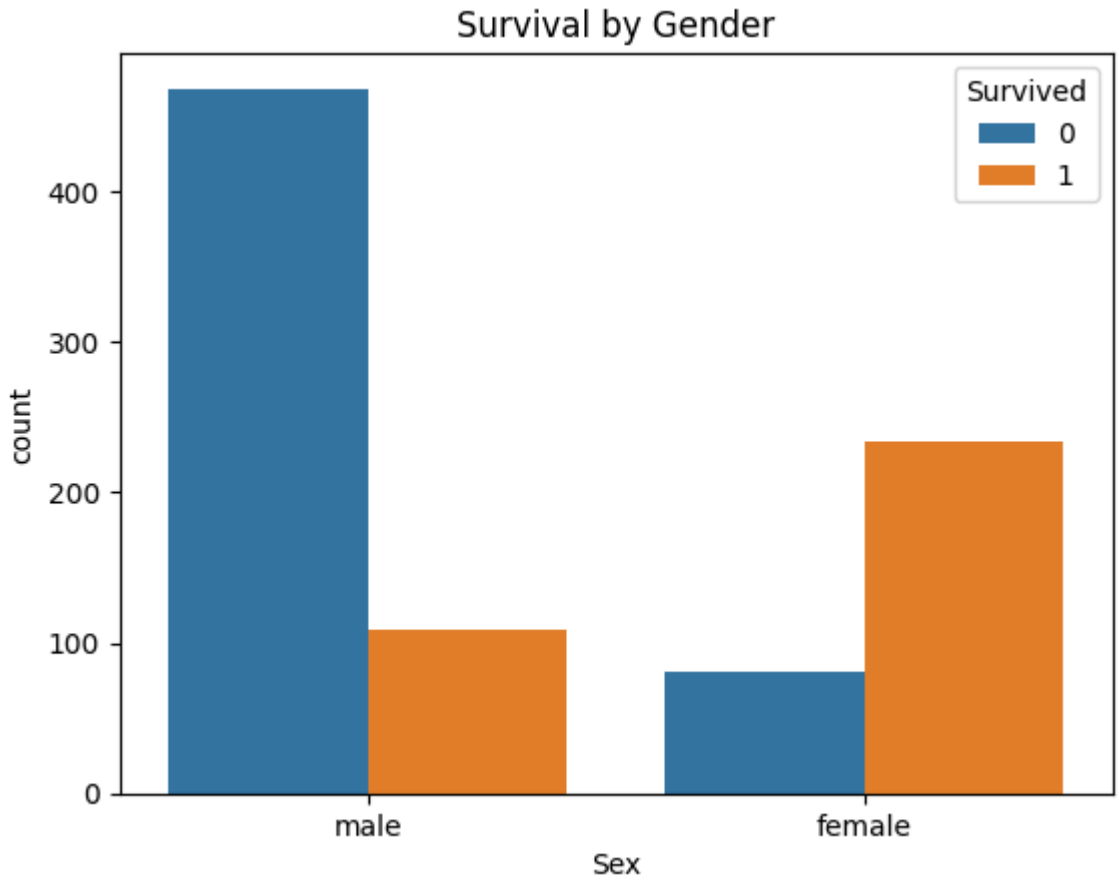
In [8]:
```
sns.histplot(df['Age'].dropna(), kde=True)
plt.title("Age Distribution")
plt.show()
```



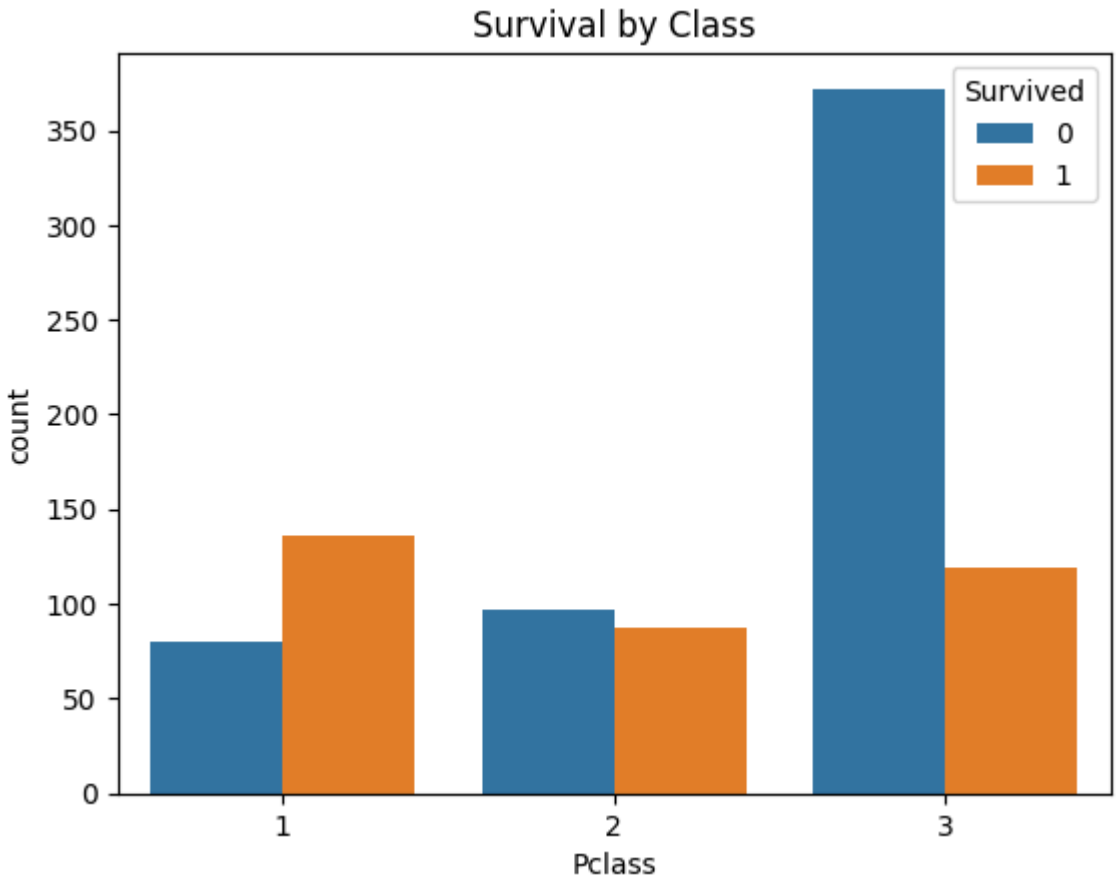
In [9]:
```
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
```



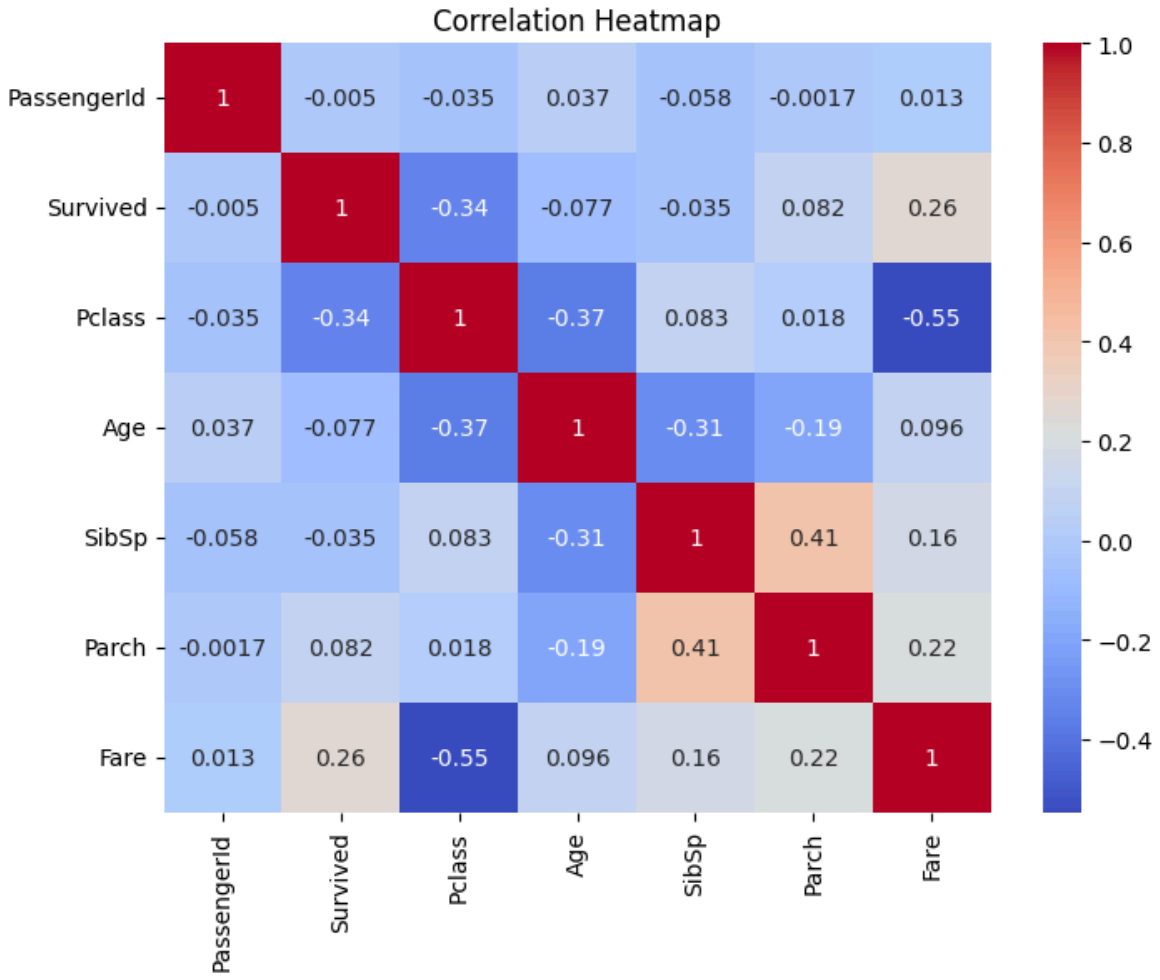
In [10]:
```
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Survival by Gender")
plt.show()
```



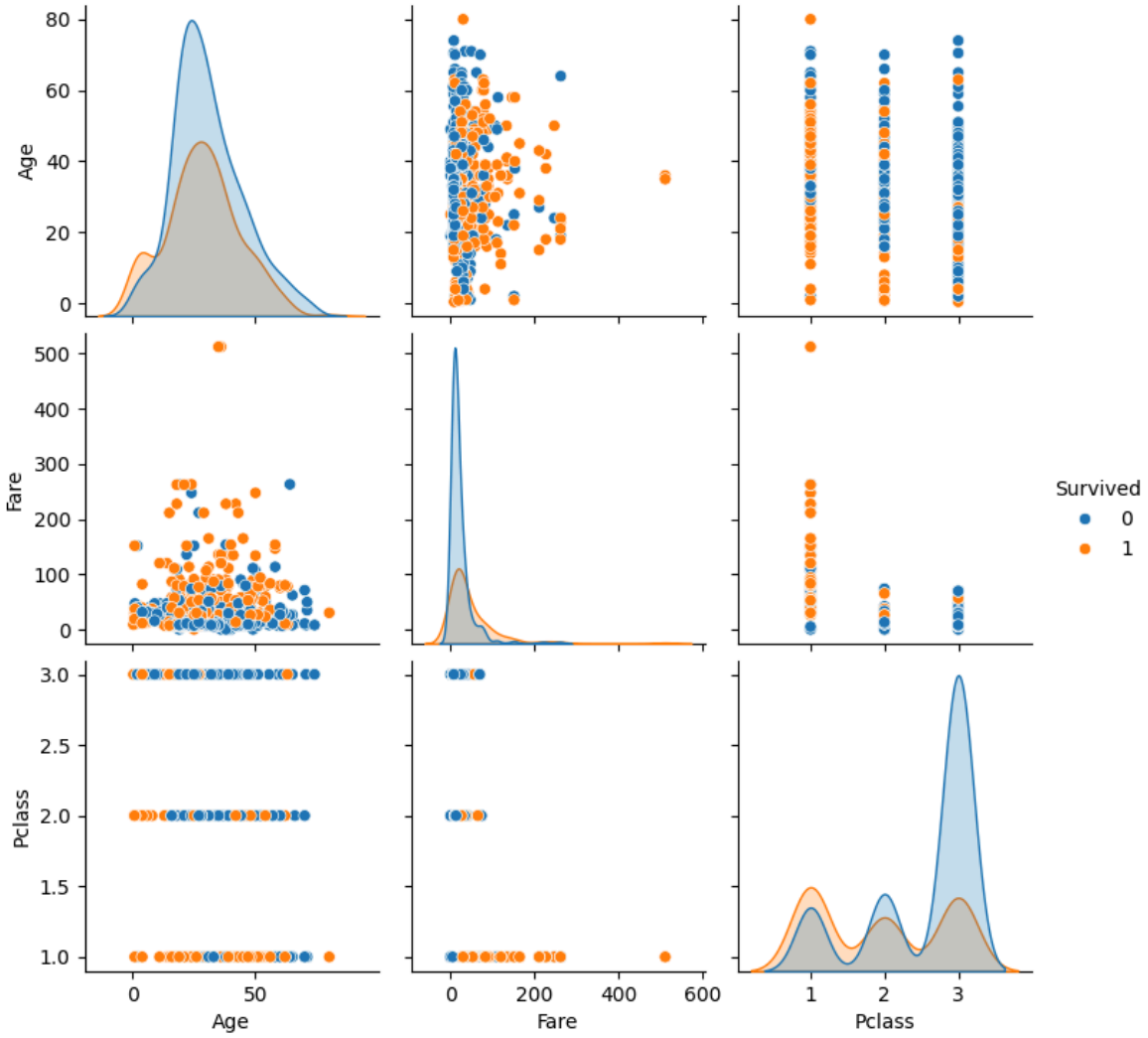
In [11]:
```
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Survival by Class")
plt.show()
```



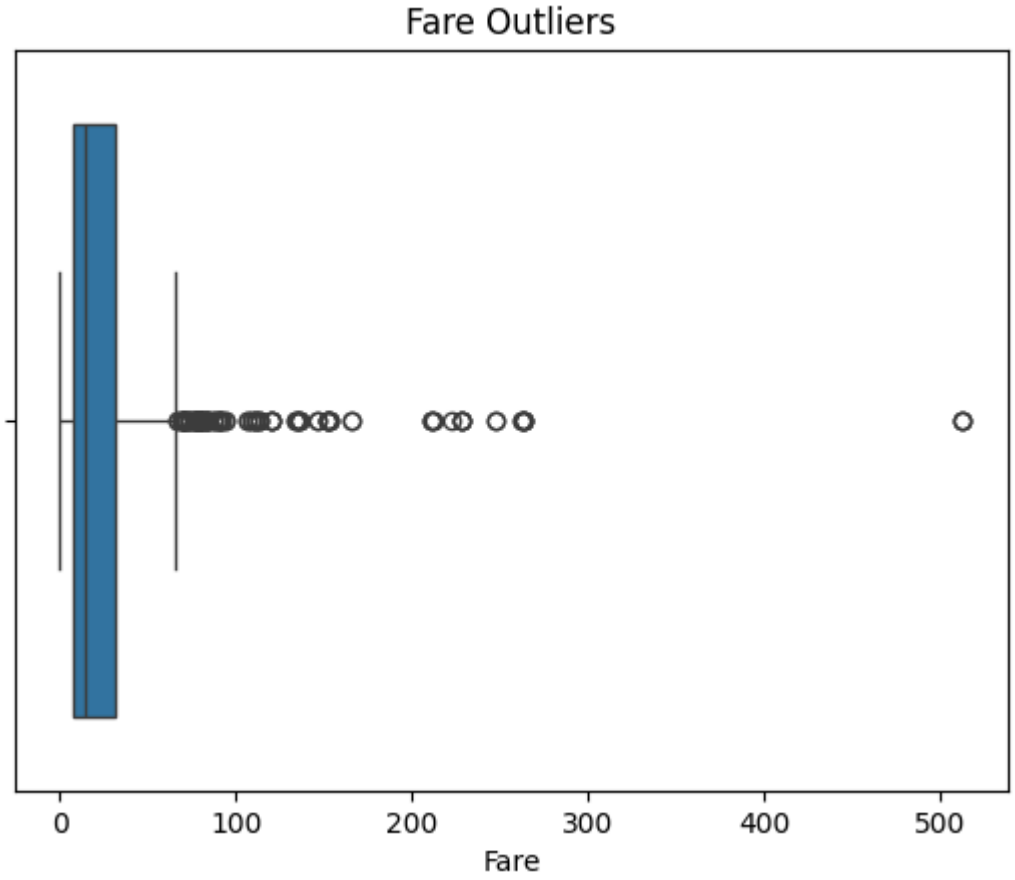
In [14]:
```
# Correlation Heatmap (numeric columns only)
plt.figure(figsize=(8,6))
numeric_df = df.select_dtypes(include=['int64','float64'])  # only numeric colum
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```



In [15]:
```
# Pairplot
sns.pairplot(df[['Survived','Age','Fare','Pclass']], hue='Survived')
plt.show()
```



In [16]:
```
sns.boxplot(x=df['Fare'])
plt.title("Fare Outliers")
plt.show()
```



In [ ]: