In [1]:
```
print("Hello")
```

```
Hello
```

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [13]:
```
import os
print(os.getcwd())
```

```
C:\Users\hp
```

In [27]:
```
df = pd.read_csv('C:/Users/hp/train.csv')
```

In [29]:
```
df.head()         # First 5 rows
df.info()         # Data types, missing values
df.describe()     # Summary stats (mean, median, etc.)
df.isnull().sum() # Missing values column-wise
df.value_counts() # Frequency of values (good for categorical variables)
```

```
<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[29]:
```
PassengerId  Survived  Pclass  Name
Sex     Age   SibSp  Parch  Ticket    Fare      Cabin  Embarked
2            1         1       Cumings, Mrs. John Bradley (Florence Briggs Thay
er)   female  38.0  1      0      PC 17599  71.2833   C85    C           1
572          1         1       Appleton, Mrs. Edward Dale (Charlotte Lamson)
female  53.0  2      0      11769     51.4792   C101   S           1
578          1         1       Silvey, Mrs. William Baird (Alice Munger)
female  39.0  1      0      13507     55.9000   E44    S           1
582          1         1       Thayer, Mrs. John Borland (Marian Longstreth Mor
ris)  female  39.0  1      1      17421     110.8833  C68    C           1
584          0         1       Ross, Mr. John Hugo
male    36.0  0      0      13049     40.1250   A10    C           1

..
328          1         2       Ball, Mrs. (Ada E Hall)
female  36.0  0      0      28551     13.0000   D      S           1
330          1         1       Hippach, Miss. Jean Gertrude
female  16.0  0      1      111361    57.9792   B18    C           1
332          0         1       Partner, Mr. Austen
male    45.5  0      0      113043    28.5000   C124   S           1
333          0         1       Graham, Mr. George Edward
male    38.0  0      1      PC 17582  153.4625  C91    S           1
890          1         1       Behr, Mr. Karl Howell
male    26.0  0      0      111369    30.0000   C148   C           1
Name: count, Length: 183, dtype: int64
```

In [37]:
```
test = pd.read_csv('C:/Users/hp/test.csv')
```

In [39]:
```
gender_submission = pd.read_csv('C:/Users/hp/gender_submission.csv')
```

In [43]:
```
df = pd.read_csv('C:/Users/hp/train.csv')
```

In [51]:
```
import pandas as pd

train = pd.read_csv('C:/Users/hp/train.csv')

train.head()
```

Out[51]:

	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 [53]:
```
train.info()


train.describe()

train['Survived'].value_counts()
train['Pclass'].value_counts()
train['Sex'].value_counts()
train['Embarked'].value_counts()

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

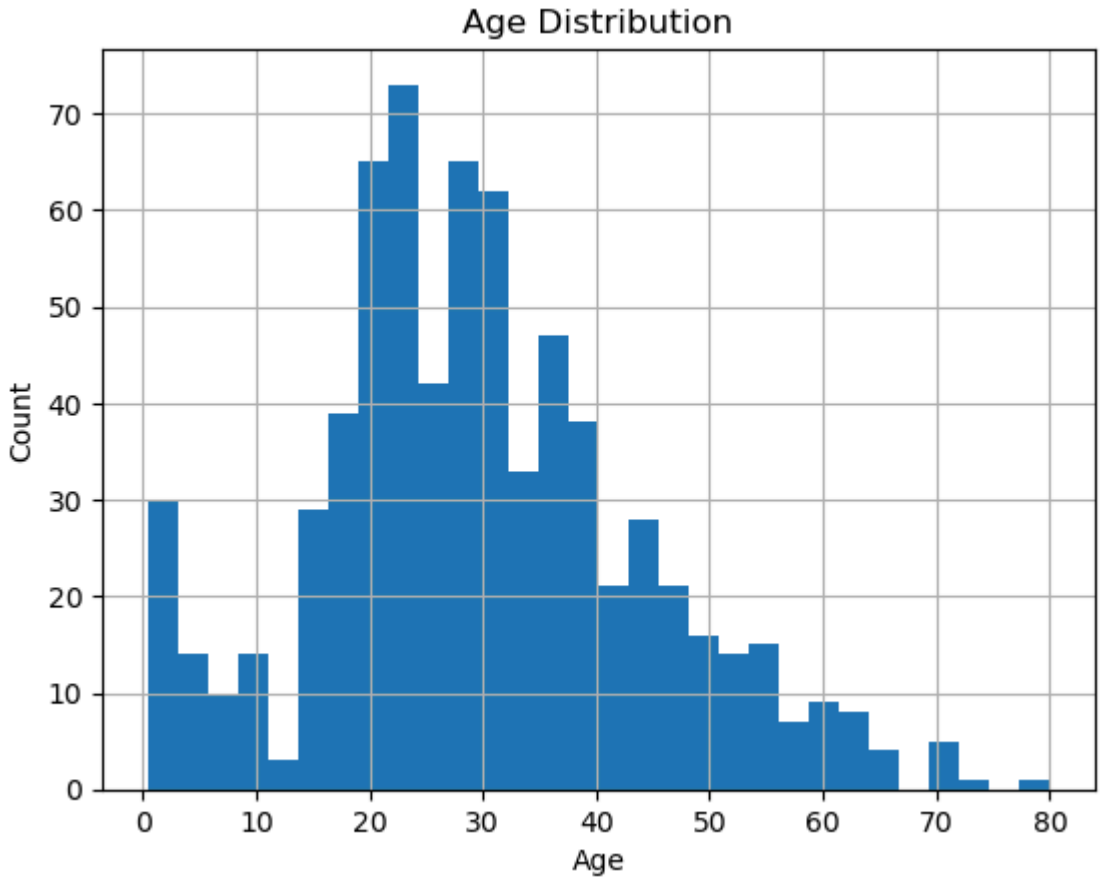
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null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

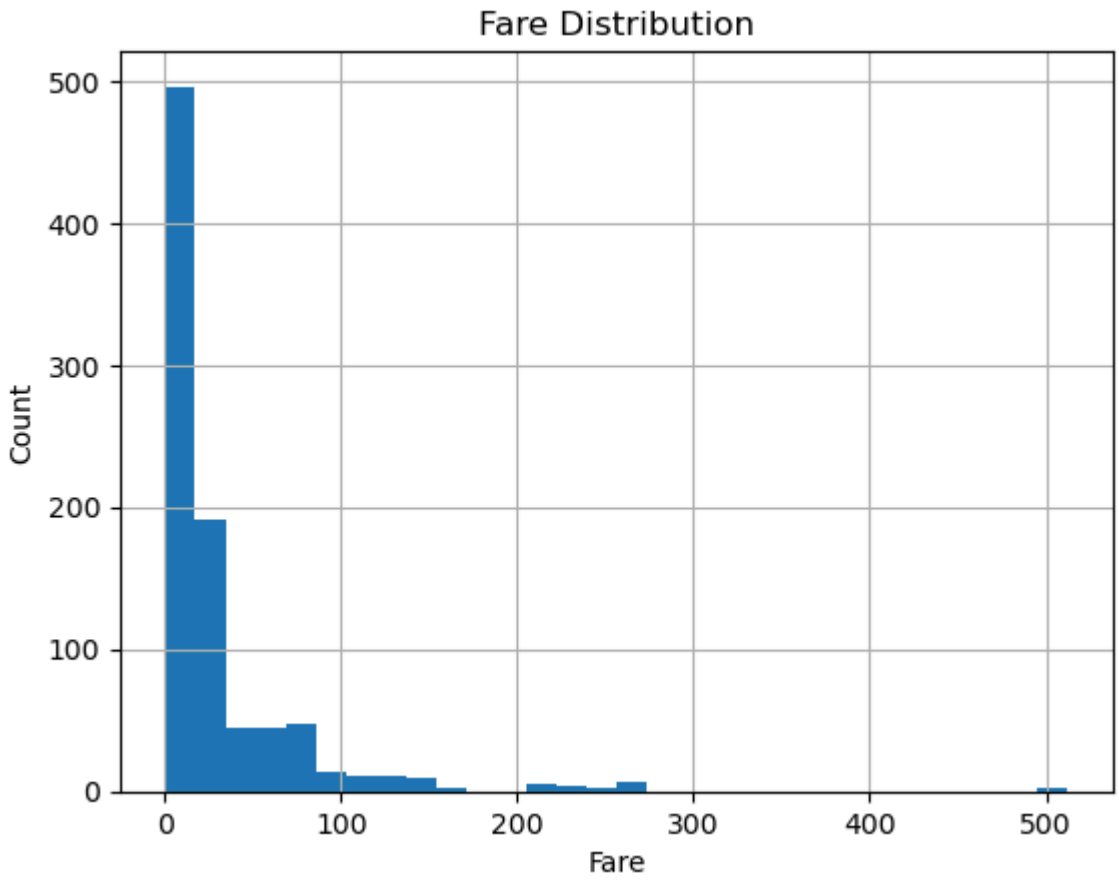
Out[53]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [55]:
```
train.isnull().sum()
```

Out[55]:
```
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
```

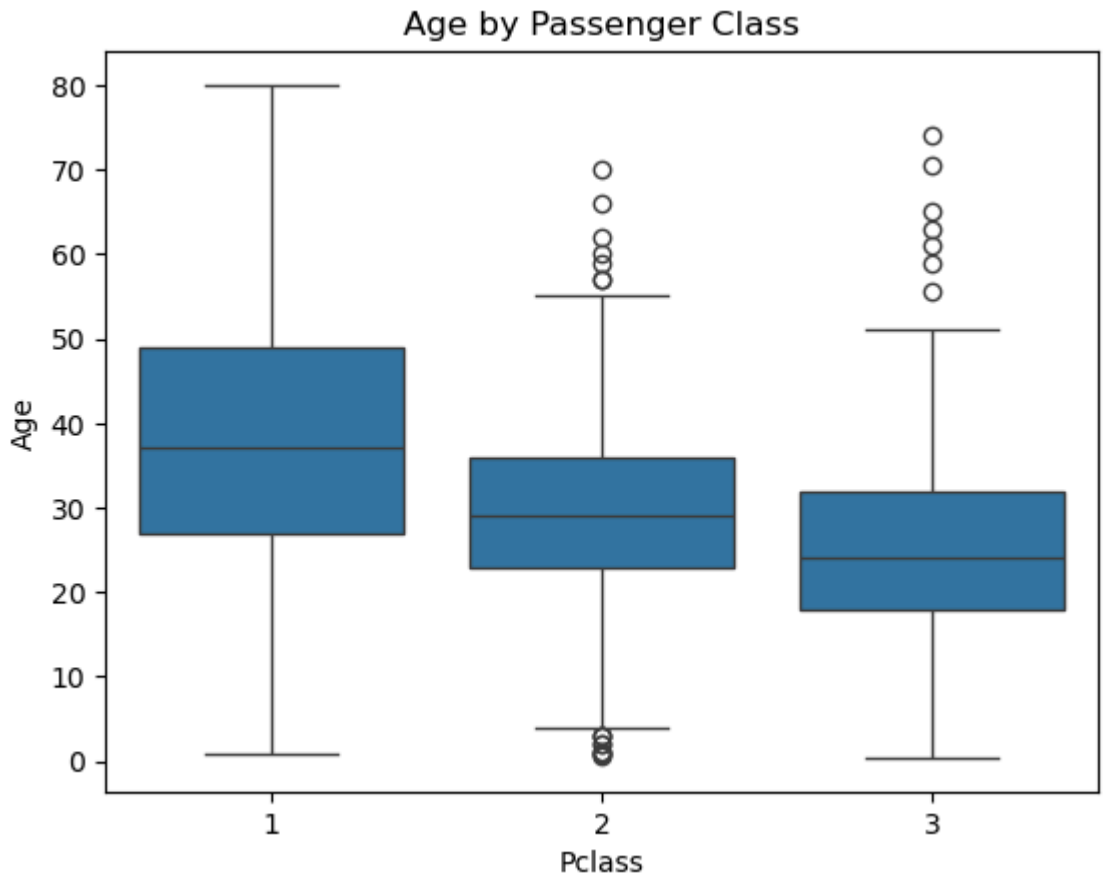
In [57]:
```
train['Age'].hist(bins=30)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()

train['Fare'].hist(bins=30)
plt.title('Fare Distribution')
plt.xlabel('Fare')
plt.ylabel('Count')
plt.show()
```

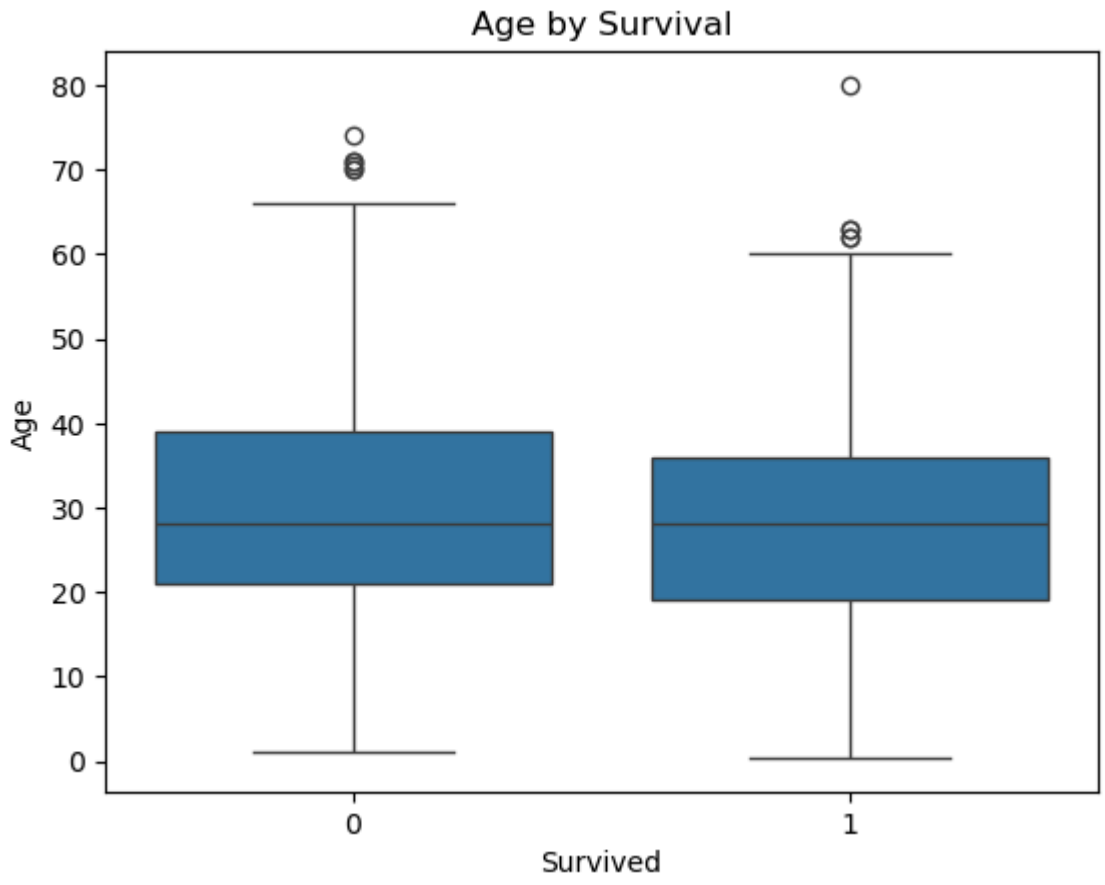




In [59]:
```
sns.boxplot(x='Pclass', y='Age', data=train)
plt.title('Age by Passenger Class')
plt.show()

sns.boxplot(x='Survived', y='Age', data=train)
```

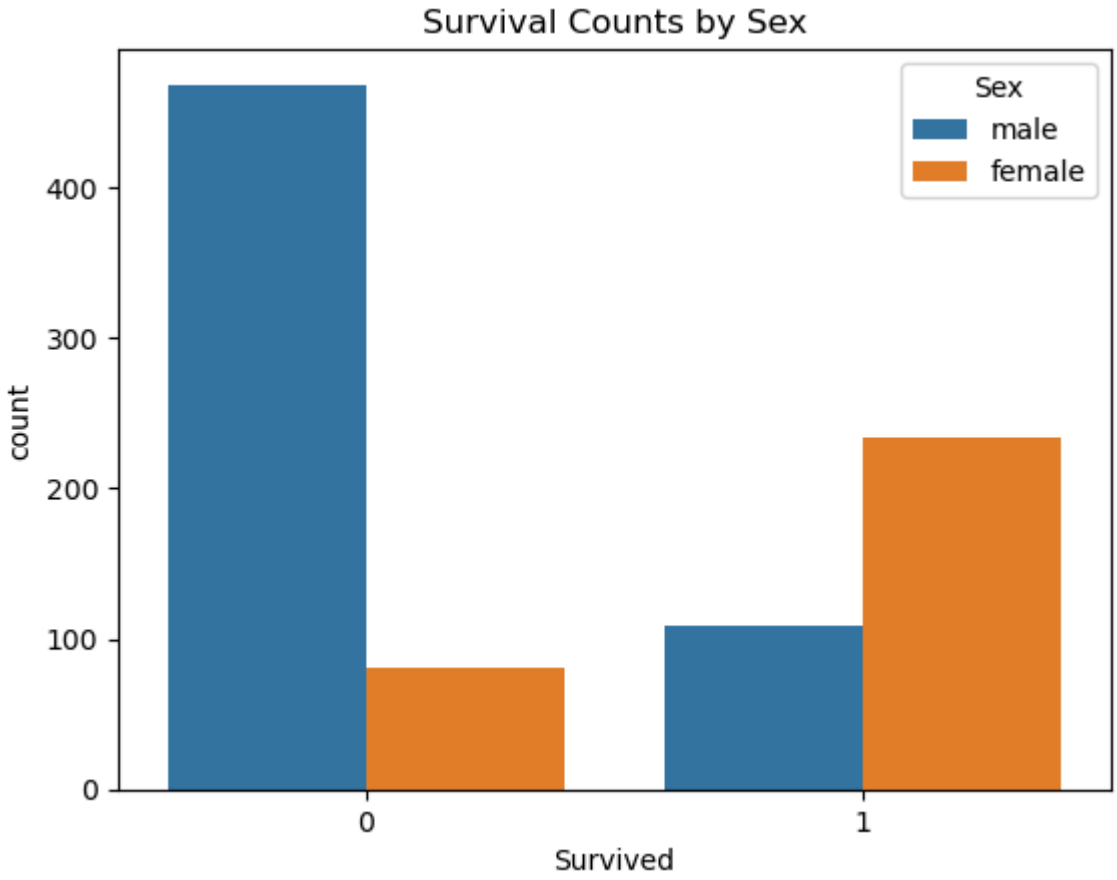
```
plt.title('Age by Survival')
plt.show()
```

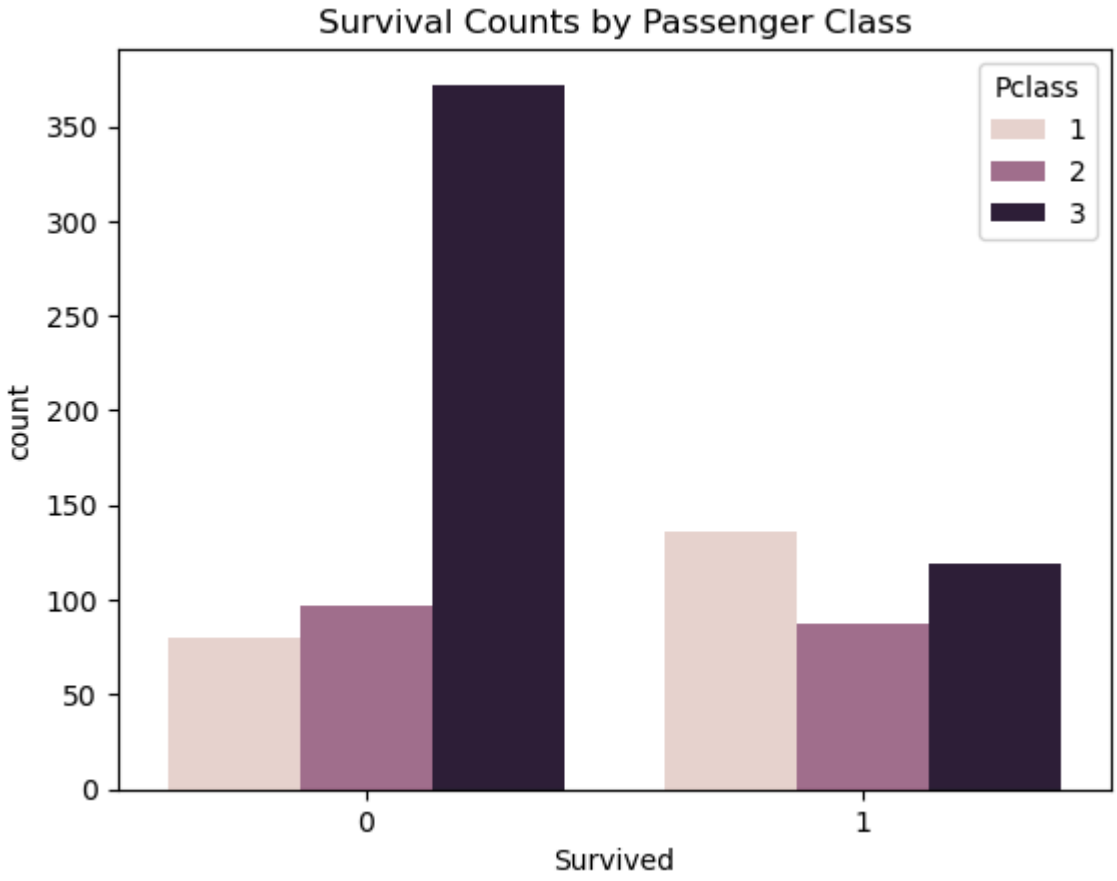




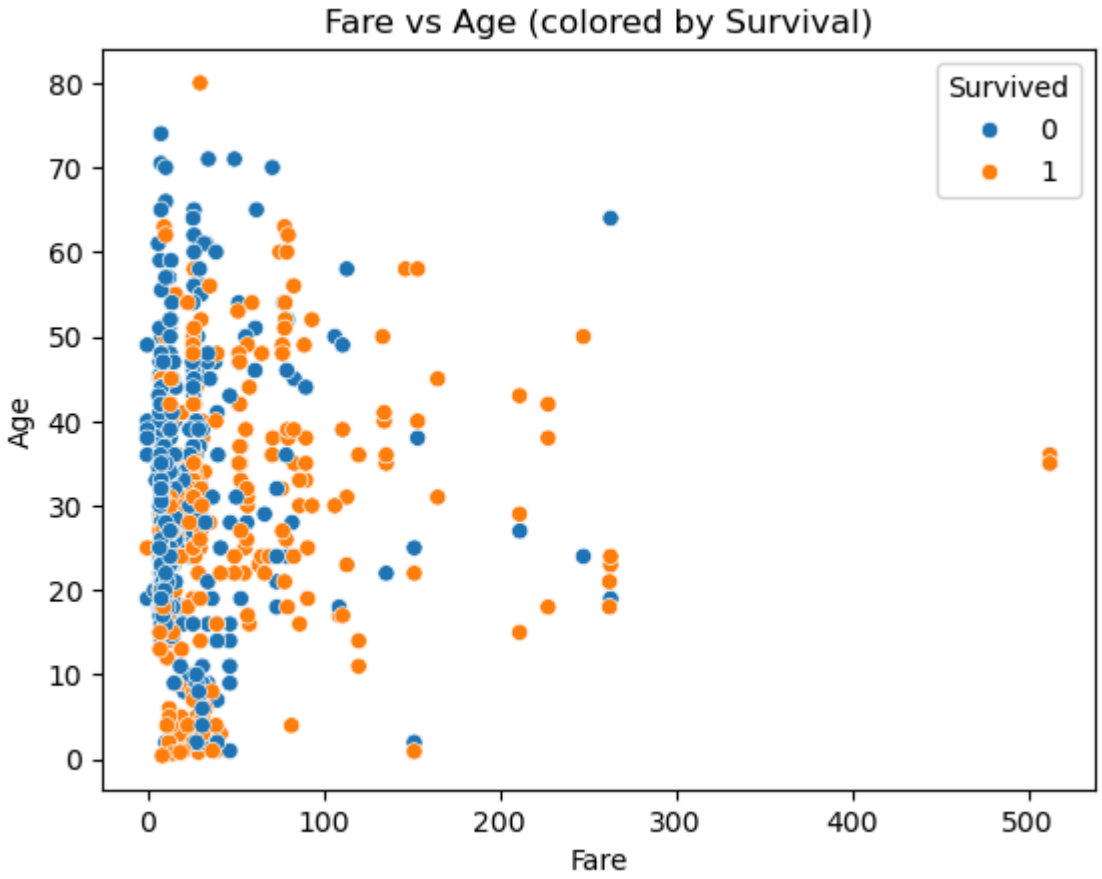
In [61]:
```
sns.countplot(x='Survived', hue='Sex', data=train)
plt.title('Survival Counts by Sex')
plt.show()
```

```
sns.countplot(x='Survived', hue='Pclass', data=train)
plt.title('Survival Counts by Passenger Class')
plt.show()
```

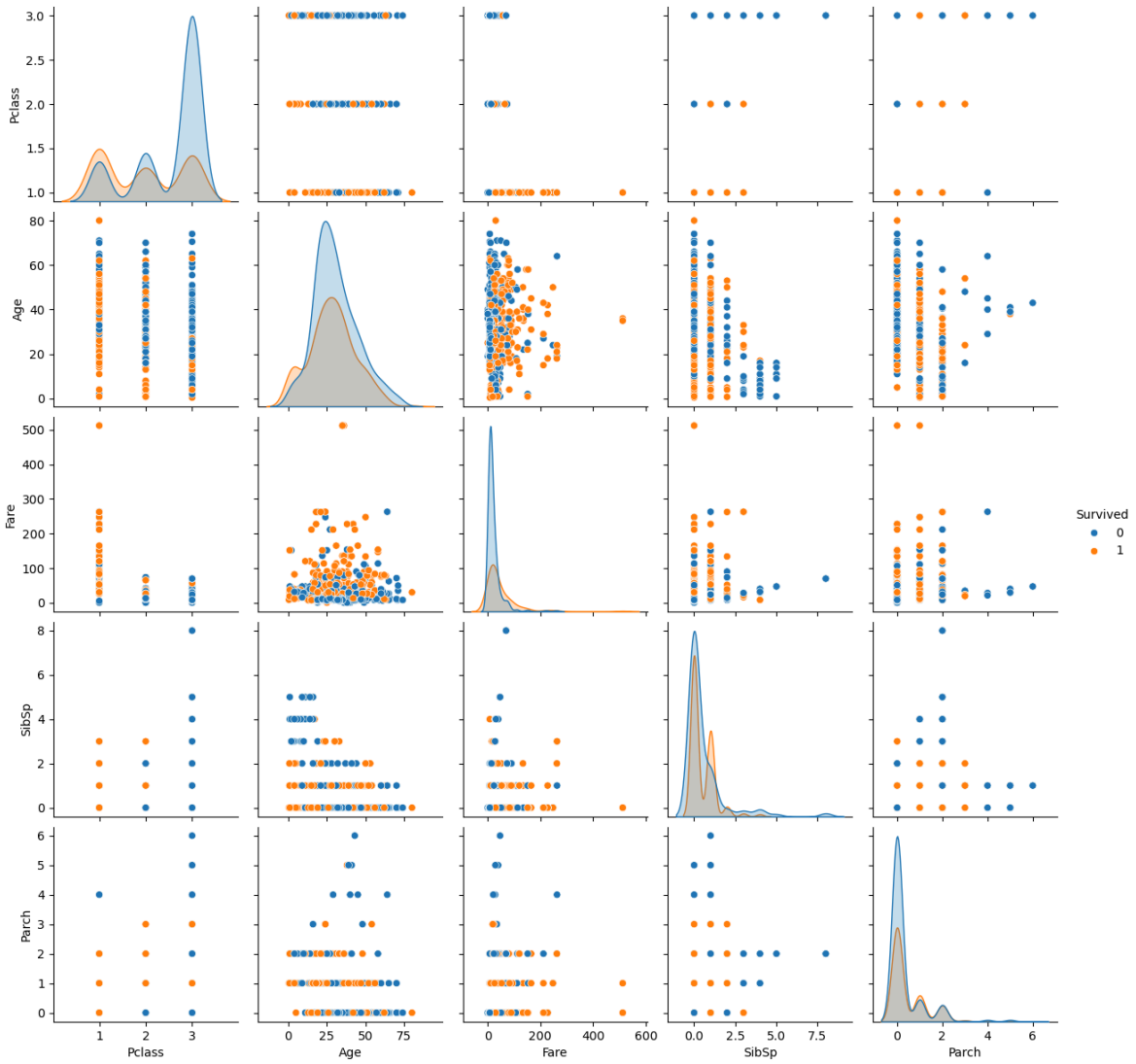




In [63]:
```
sns.scatterplot(x='Fare', y='Age', hue='Survived', data=train)
plt.title('Fare vs Age (colored by Survival)')
plt.show()
```



In [67]:
```
sns.pairplot(train[['Survived', 'Pclass', 'Age', 'Fare', 'SibSp', 'Parch']], hue
plt.show()
```

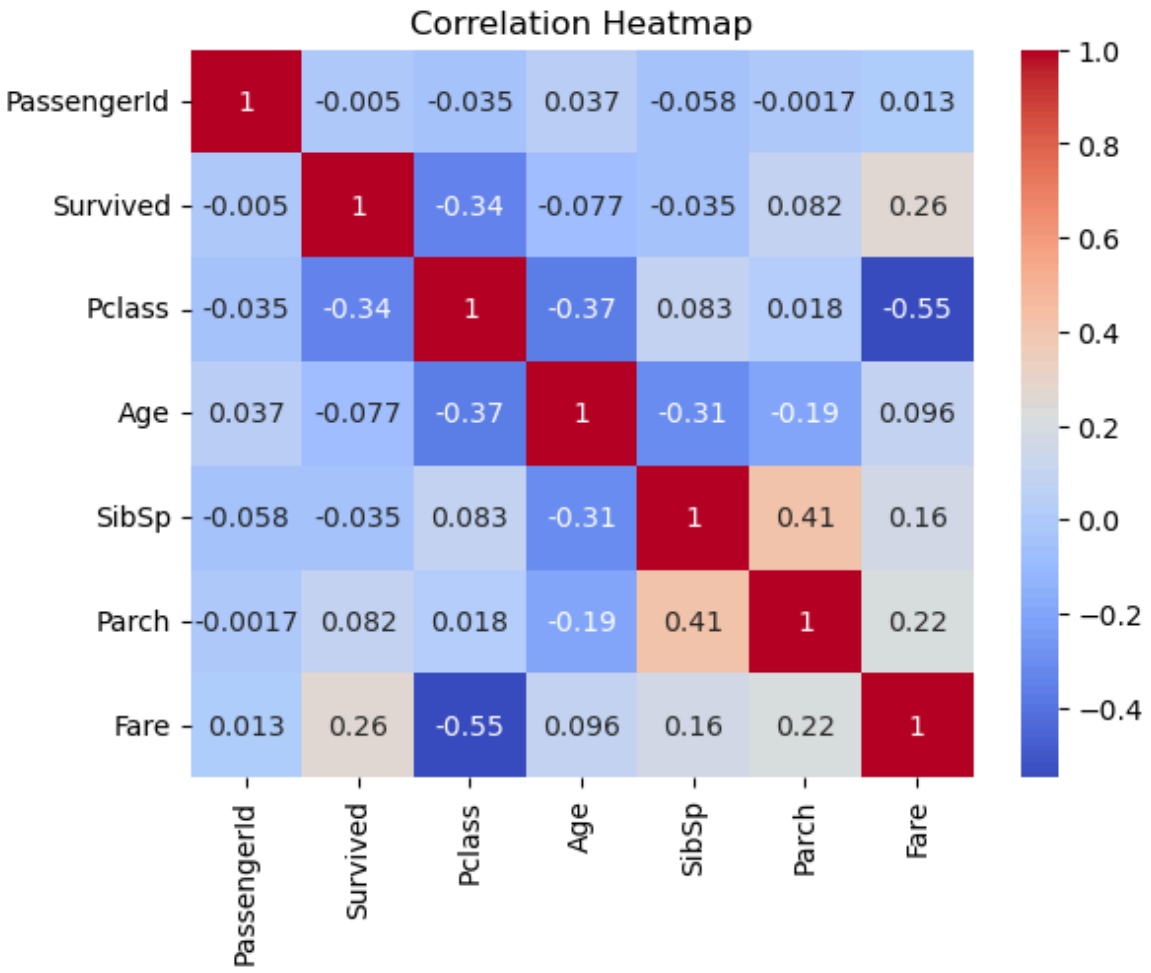


In [69]:
```
correlation = train.corr()
sns.heatmap(correlation, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[69], line 1
----> 1 correlation = train.corr()
      2 sns.heatmap(correlation, annot=True, cmap='coolwarm')
      3 plt.title('Correlation Heatmap')

File D:\c++\Anaconda\Lib\site-packages\pandas\core\frame.py:11049, in DataFrame.c
orr(self, method, min_periods, numeric_only)
  11047 cols = data.columns
  11048 idx = cols.copy()
> 11049 mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  11051 if method == "pearson":
  11052     correl = libalgos.nancorr(mat, minp=min_periods)

File D:\c++\Anaconda\Lib\site-packages\pandas\core\frame.py:1993, in DataFrame.to
_numpy(self, dtype, copy, na_value)
   1991 if dtype is not None:
   1992     dtype = np.dtype(dtype)
-> 1993 result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)
   1994 if result.dtype is not dtype:
   1995     result = np.asarray(result, dtype=dtype)

File D:\c++\Anaconda\Lib\site-packages\pandas\core\internals\managers.py:1694, in
BlockManager.as_array(self, dtype, copy, na_value)
   1692         arr.flags.writeable = False
   1693 else:
-> 1694     arr = self._interleave(dtype=dtype, na_value=na_value)
   1695     # The underlying data was copied within _interleave, so no need
   1696     # to further copy if copy=True or setting na_value
   1698 if na_value is lib.no_default:

File D:\c++\Anaconda\Lib\site-packages\pandas\core\internals\managers.py:1753, in
BlockManager._interleave(self, dtype, na_value)
   1751     else:
   1752         arr = blk.get_values(dtype)
-> 1753     result[rl.indexer] = arr
   1754     itemmask[rl.indexer] = 1
   1756 if not itemmask.all():

ValueError: could not convert string to float: 'Braund, Mr. Owen Harris'
```

In [71]:

```
numeric_train = train.select_dtypes(include=['number'])


correlation = numeric_train.corr()


sns.heatmap(correlation, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

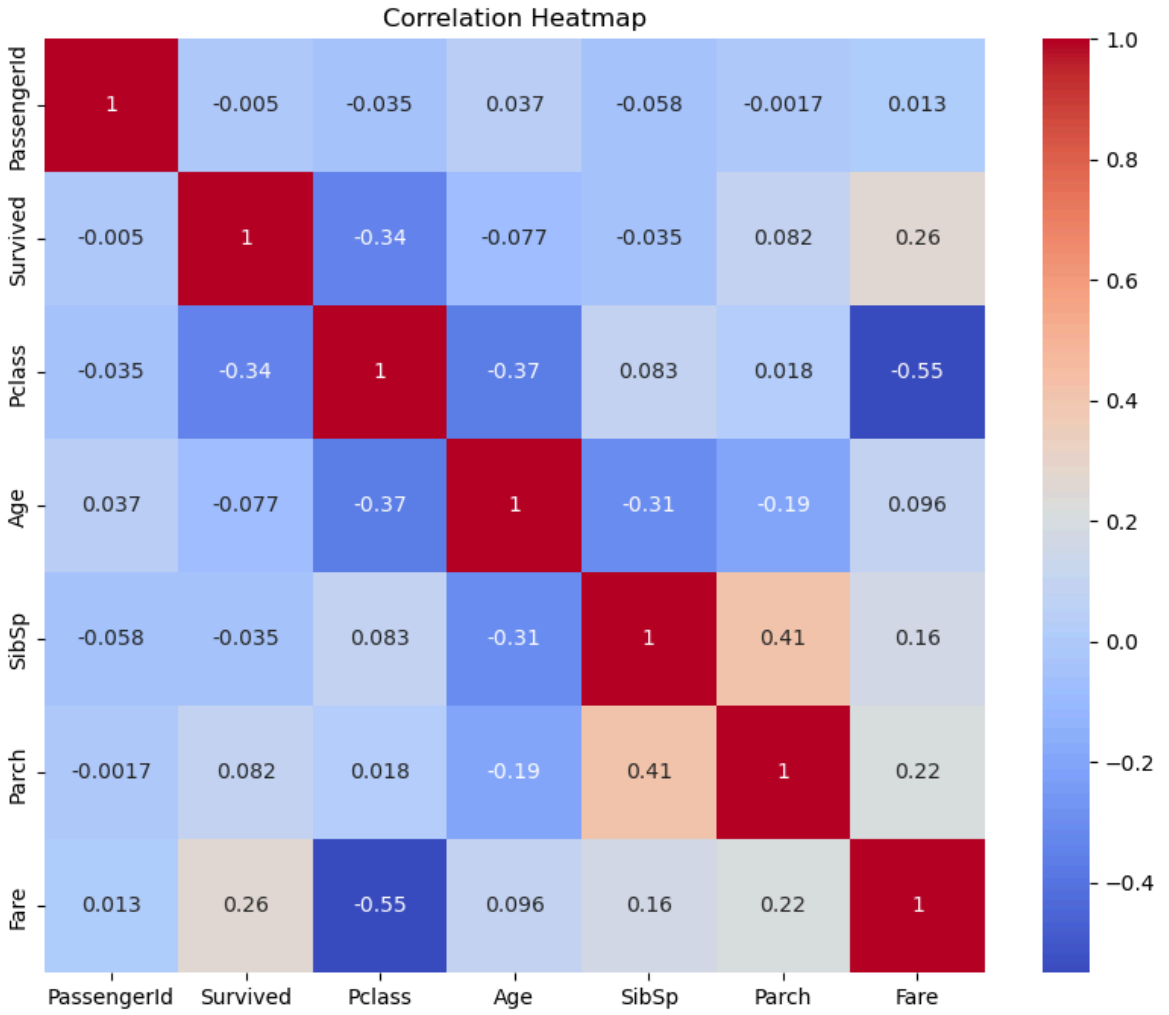


In [73]:
```
train.dtypes
```

Out[73]:
```
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 [75]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


numeric_train = train.select_dtypes(include=['number'])

# Correlation
correlation = numeric_train.corr()

# Plot
plt.figure(figsize=(10,8))
sns.heatmap(correlation, annot=True, cmap='coolwarm')
```

```
plt.title('Correlation Heatmap')
plt.show()
```



In [ ]: