In [14]:

```python
# Exploratory Data Analysis (EDA) on Titanic CSV Dataset

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Set visual style
sns.set(style="whitegrid")

# --- Step 1: Read CSV and Create DataFrame ---
df = pd.read_csv('titanic_sample.csv')  # Make sure the file is in the same directory
```

In [9]:

```python
# --- Step 2: Preview Dataset ---
print(" First 5 Rows:")
print(df.head())

print("\n✅  Dataset Shape:", df.shape)
print("\n Column Names:", df.columns.tolist())

# --- Step 3: Summary Info ---
print("\n Data Types:")
print(df.dtypes)

print("\n Summary Statistics:")
print(df.describe(include='all'))

# --- Step 4: Check Missing Values ---
print("\n! Missing Values:")
print(df.isnull().sum())
```

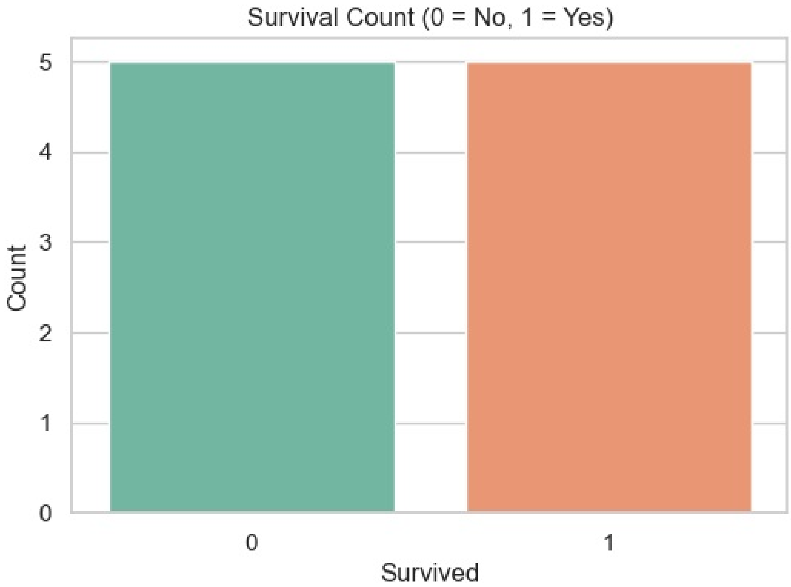
```
 First 5 Rows:
   survived  pclass     sex   age  sibsp  parch     fare embarked  class  \
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True

✅  Dataset Shape: (10, 15)

 Column Names: ['survived', 'pclass', 'sex', 'age', 'sibsp', 'parch', 'fare', 'embarked', 'class', 'who', 'adult
_male', 'deck', 'embark_town', 'alive', 'alone']

 Data Types:
survived         int64
pclass           int64
sex             object
age            float64
sibsp            int64
parch            int64
fare           float64
embarked        object
class           object
who             object
adult_male        bool
deck            object
embark_town     object
alive           object
alone             bool
dtype: object

 Summary Statistics:
         survived     pclass   sex        age      sibsp      parch  \
count   10.000000  10.000000    10   9.000000  10.000000  10.000000
unique        NaN        NaN     2        NaN        NaN        NaN
top           NaN        NaN  male        NaN        NaN        NaN
freq          NaN        NaN     5        NaN        NaN        NaN
mean     0.500000   2.300000   NaN  28.111111   0.700000   0.300000
std      0.527046   0.948683   NaN  14.945271   0.948683   0.674949
min      0.000000   1.000000   NaN   2.000000   0.000000   0.000000
25%      0.000000   1.250000   NaN  22.000000   0.000000   0.000000
50%      0.500000   3.000000   NaN  27.000000   0.500000   0.000000
75%      1.000000   3.000000   NaN  35.000000   1.000000   0.000000
max      1.000000   3.000000   NaN  54.000000   3.000000   2.000000
```

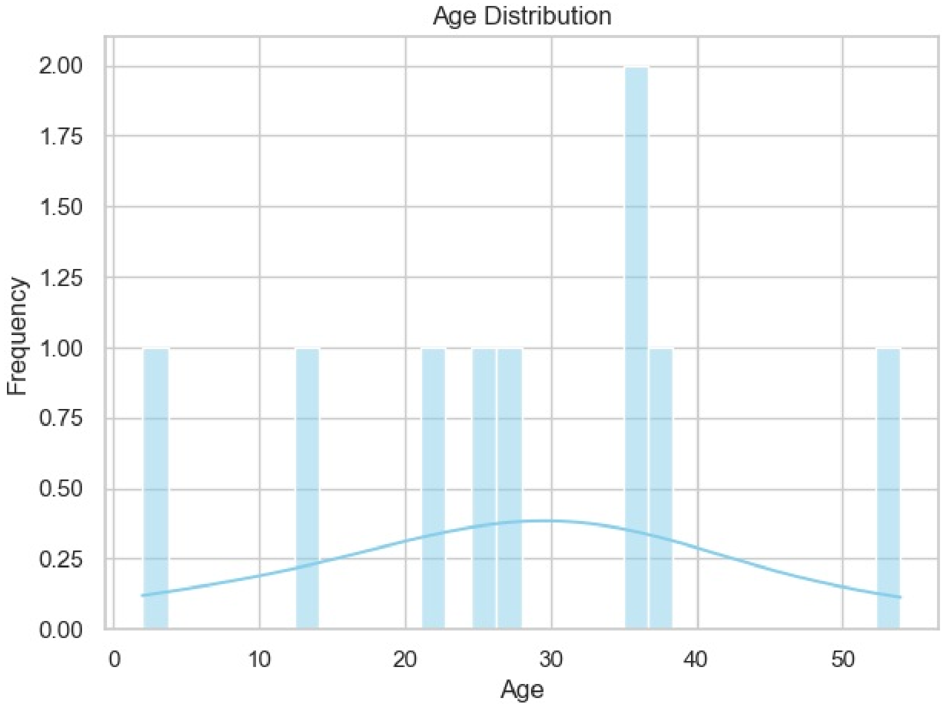
```
             fare embarked  class  who adult_male deck  embark_town alive  \
count   10.000000       10     10   10         10    3           10    10   
unique        NaN        3      3    3          2    2            3     2   
top           NaN        S  Third  man      False    C  Southampton    no   
freq          NaN        7      6    4          6    2            7     5   
mean    27.020820      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
std     23.601938      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
min      7.250000      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
25%      8.152075      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
50%     16.104150      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
75%     46.414575      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
max     71.283300      NaN    NaN  NaN        NaN  NaN          NaN   NaN   

        alone  
count      10  
unique      2  
top     False  
freq        6  
mean      NaN  
std       NaN  
min       NaN  
25%       NaN  
50%       NaN  
75%       NaN  
max       NaN  

! Missing Values:
survived       0
pclass         0
sex            0
age            1
sibsp          0
parch          0
fare           0
embarked       0
class          0
who            0
adult_male     0
deck           7
embark_town    0
alive          0
alone          0
dtype: int64
```

In [11]:

```python
# --- Step 5: Univariate Analysis ---

# Survival Count
plt.figure(figsize=(6, 4))
sns.countplot(x='survived', data=df, palette='Set2')
plt.title('Survival Count (0 = No, 1 = Yes)')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()

# Age Distribution
plt.figure(figsize=(7, 5))
sns.histplot(df['age'], bins=30, kde=True, color='skyblue')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

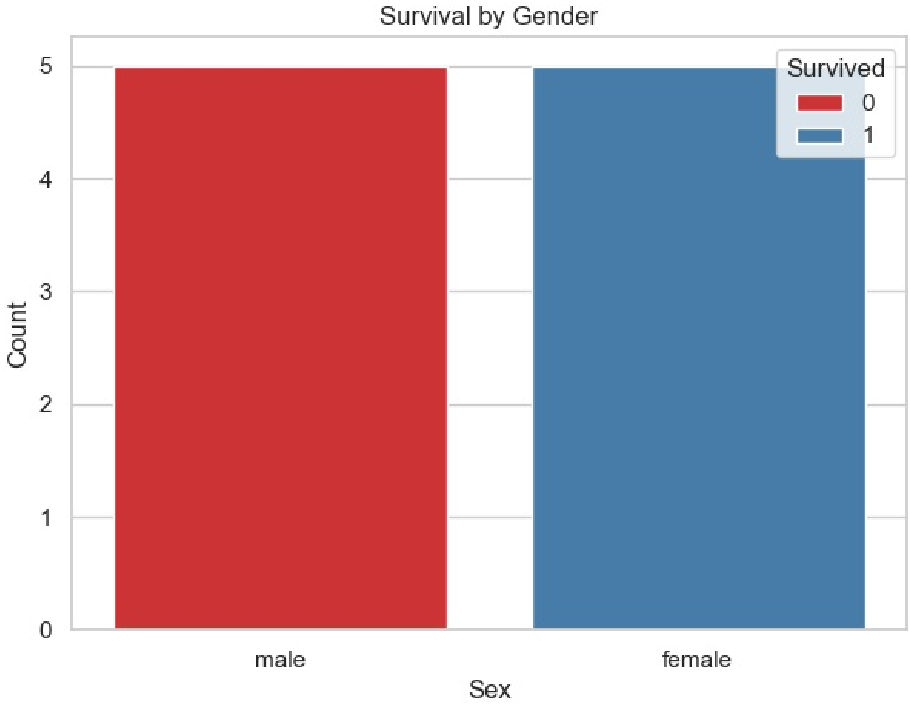
```
C:\Users\ELCOT\AppData\Local\Temp\ipykernel_12880\42462311.py:5: FutureWarning: 

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable
to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='survived', data=df, palette='Set2')
```



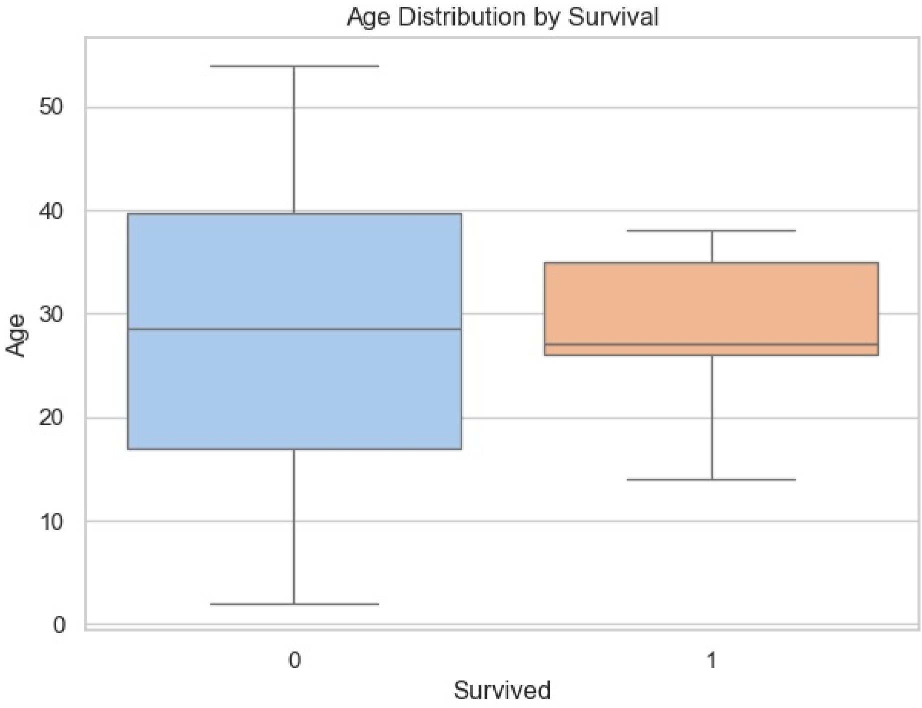


In [12]:

```
# --- Step 6: Bivariate Analysis ---

# Survival by Gender
plt.figure(figsize=(7, 5))
sns.countplot(x='sex', hue='survived', data=df, palette='Set1')
plt.title('Survival by Gender')
plt.xlabel('Sex')
plt.ylabel('Count')
plt.legend(title='Survived')
plt.show()

# Age vs Survival Boxplot
plt.figure(figsize=(7, 5))
sns.boxplot(x='survived', y='age', data=df, palette='pastel')
plt.title('Age Distribution by Survival')
plt.xlabel('Survived')
plt.ylabel('Age')
plt.show()
```

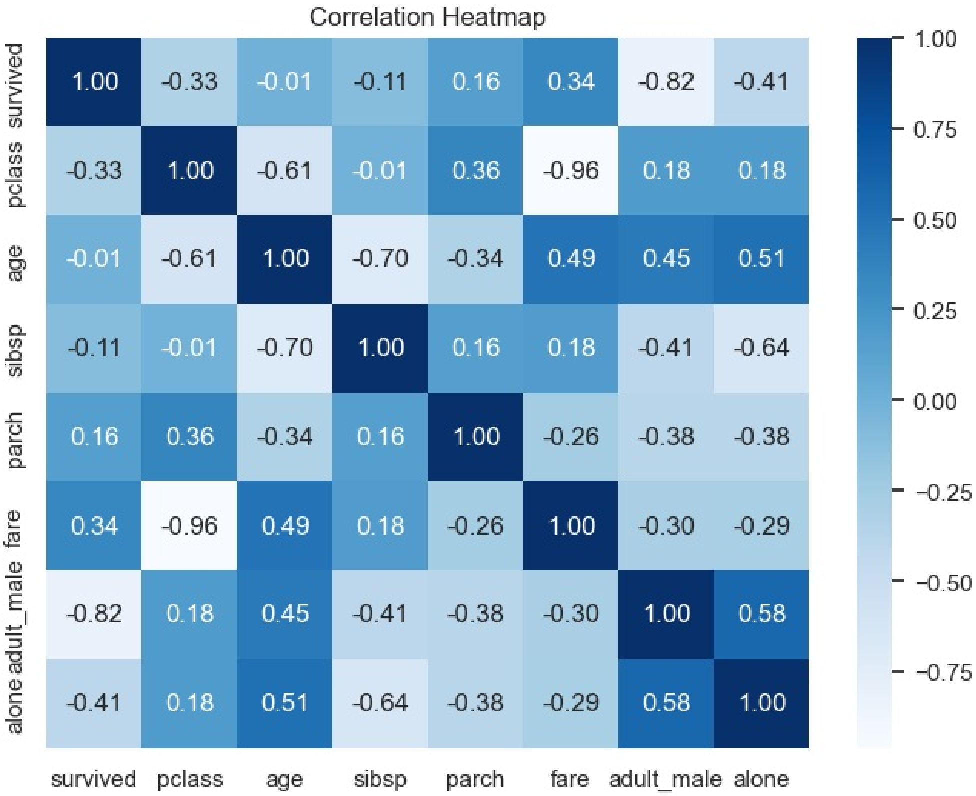


```
C:\Users\ELCOT\AppData\Local\Temp\ipykernel_12880\1820436489.py:14: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable
to `hue` and set `legend=False` for the same effect.

  sns.boxplot(x='survived', y='age', data=df, palette='pastel')
```



In [13]:

```
# --- Step 7: Correlation Heatmap ---

plt.figure(figsize=(8, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='Blues', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js