Assignment 8 - Data Visualization I (Data Science) ¶

Atharva Taras (TE A - 73)

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.

Dataset - Kaggle Titanic (https://www.kaggle.com/competitions/titanic/data?select=train.csv)

In [1]:

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

In [2]:

```
data = pd.read_csv('titanic_train.csv')
```

In [3]:

```
data.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

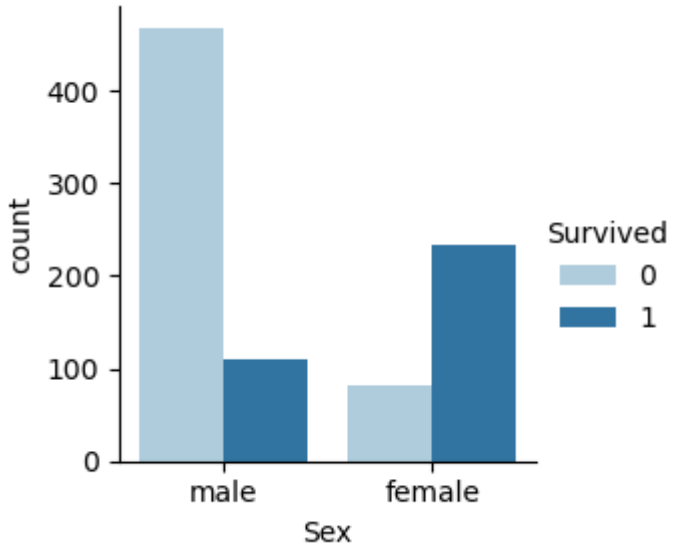
No. of people survived by gender

In [4]:

```
sns.set_palette(palette="Paired")
sns.catplot(x="Sex", hue="Survived", kind="count", data=data, height=3)
```

Out[4]:

```
<seaborn.axisgrid.FacetGrid at 0x2a69fad9730>
```



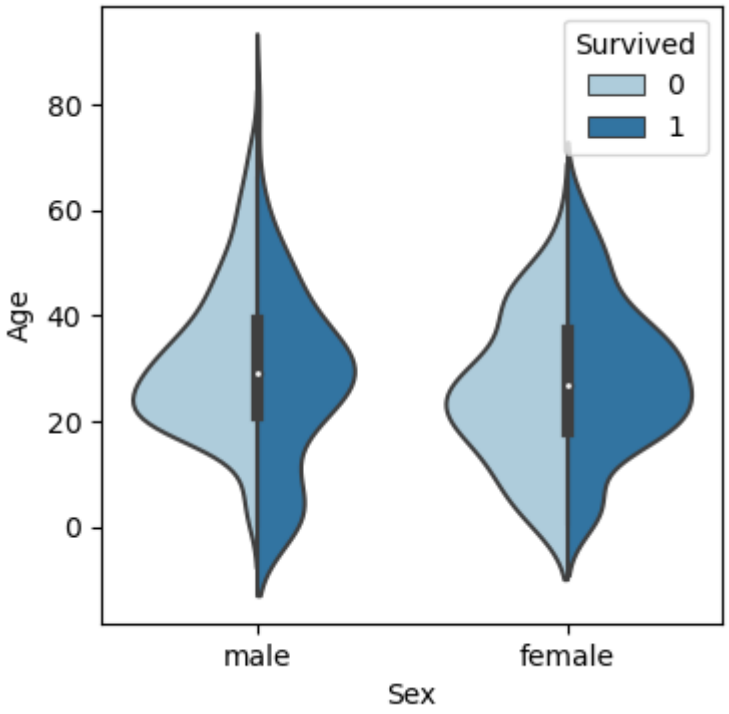
Passenger age distribution wrt survival

In [5]:

```
plt.figure(figsize=(4, 4))
sns.violinplot(x="Sex", y="Age", hue="Survived", data=data, split=True)
```

Out[5]:

```
<AxesSubplot: xlabel='Sex', ylabel='Age'>
```



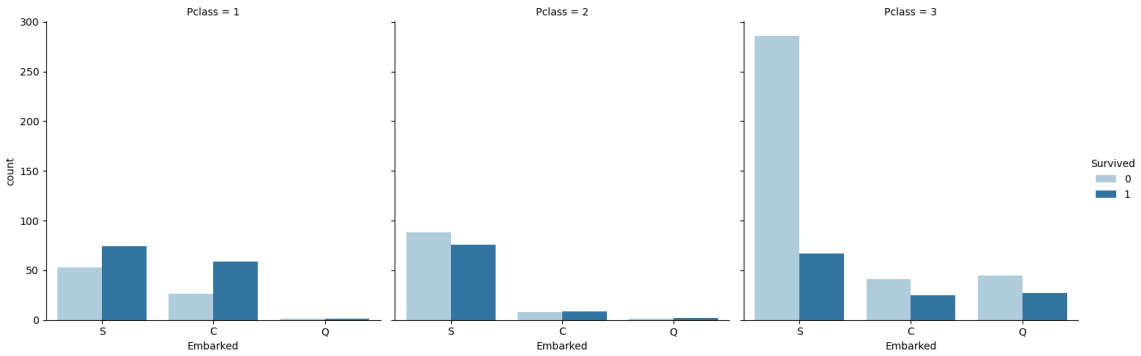
No. of passengers who survived based on passenger class

In [6]:

```
1  sns.catplot(x='Embarked', hue='Survived', kind='count', col='Pclass', data=data)
```

Out[6]:

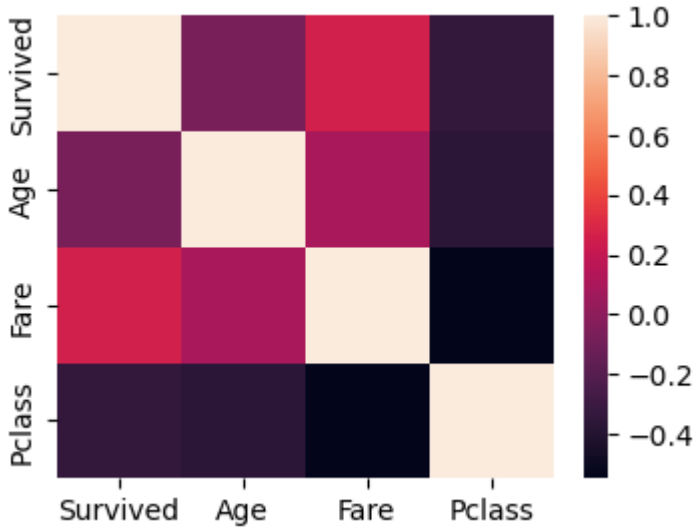
```
<seaborn.axisgrid.FacetGrid at 0x2a69fd17d60>
```



Correlation Heatmap

In [7]:

```
1  plt.figure(figsize=(4, 3))
2  dataplot=sns.heatmap(data[['Survived', 'Age', 'Fare', 'Pclass']].corr())
```



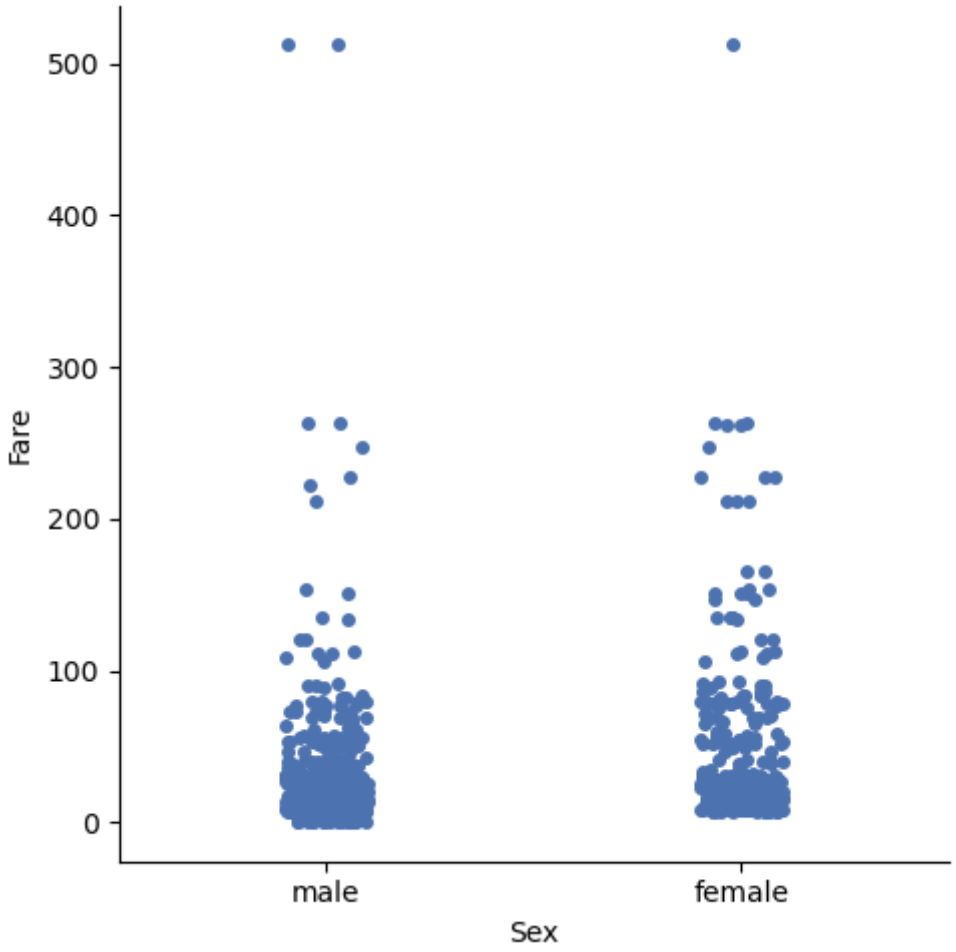
Fares paid by male and female passengers

In [8]:

```
1 sns.set_palette(palette="deep")
2 sns.catplot(data=data, x='Sex', y='Fare', height=5)
```

Out[8]:

```
<seaborn.axisgrid.FacetGrid at 0x2a6a0171220>
```



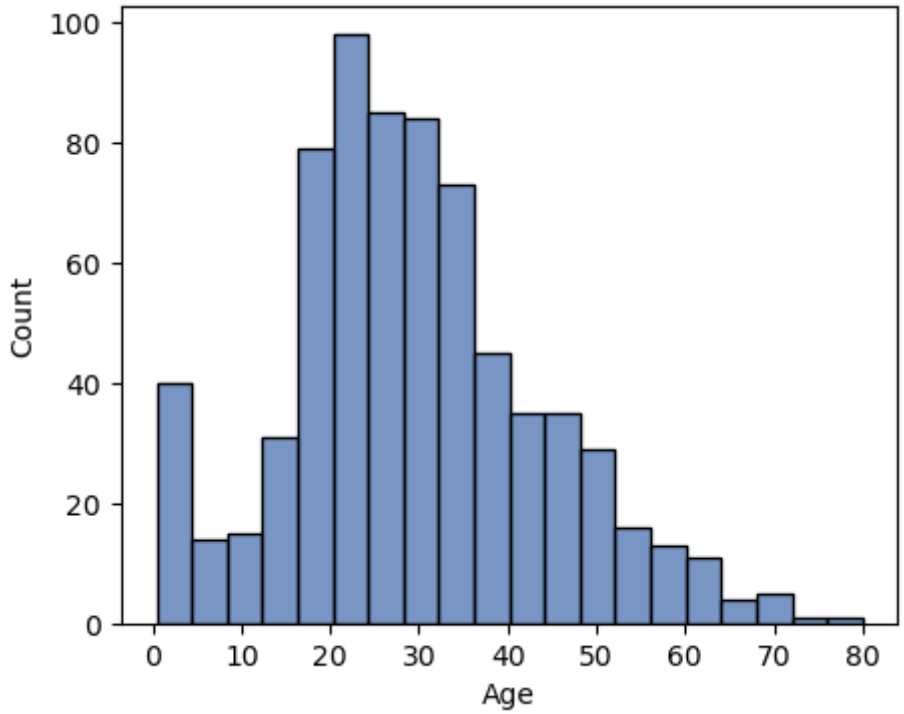
Age distribution

In [9]:

```
plt.figure(figsize=(5, 4))
sns.histplot(data['Age'])
```

Out[9]:

```
<AxesSubplot: xlabel='Age', ylabel='Count'>
```



2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

In [10]:

```
plt.figure(figsize=(10, 4))
sns.histplot(data['Fare'], bins=25)
```

Out[10]:

```
<AxesSubplot: xlabel='Fare', ylabel='Count'>
```