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

```
df = pd.read_csv("/content/titanic.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 13 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  156 non-null    int64
 1   Survived     156 non-null    int64
 2   Pclass       156 non-null    int64
 3   Lname        156 non-null    object
 4   Name         156 non-null    object
 5   Sex          156 non-null    object
 6   Age          126 non-null    float64
 7   SibSp        156 non-null    int64
 8   Parch        156 non-null    int64
 9   Ticket       156 non-null    object
 10  Fare         156 non-null    float64
 11  Cabin
 12  Embarked
dtypes: float64(2), int64(5), object(6)
memory usage: 16.0+ KB
```



```
df.shape
```

```
(156, 13)
```

```
df.columns
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Lname', 'Name', 'Sex', 'Age',
       'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```
df.size
```

```
2028
```

```
df.dtypes
```

```
PassengerId      int64
Survived         int64
Pclass           int64
Lname           object
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

```
df.head()
```

	PassengerId	Survived	Pclass	Lname	Name	Sex	Age	SibSp	Parch	Ti
0	1	0	3	Braund	Mr. Owen Harris	male	22.0	1	0	A/5 2
1	2	1	1	Cumings	Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 1
2	3	1	3	Heikkinen	Miss. Laina	female	26.0	0	0	STON 310
3	4	1	1	Futrelle	Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	11

Next steps: Generate code with `df` View recommended plots

```
df.tail()
```

	PassengerId	Survived	Pclass	Lname	Name	Sex	Age	SibSp	Parch	Tick
151	152	1	1	Pears	Mrs. Thomas (Edith Wearne)	female	22.0	1	0	1137
152	153	0	3	Meo	Mr. Alfonzo	male	55.5	0	0	A 112
153	154	0	3	van Billiard	Mr. Austin Blyler	male	40.5	0	2	A 8
154	155	0	3	Olsen	Mr. Ole Martin	male	NaN	0	0	2653


```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	F
count	156.000000	156.000000	156.000000	126.000000	156.000000	156.000000	156.000
mean	78.500000	0.346154	2.423077	28.141508	0.615385	0.397436	28.109
std	45.177428	0.477275	0.795459	14.613880	1.056235	0.870146	39.401
min	1.000000	0.000000	1.000000	0.830000	0.000000	0.000000	6.750
25%	39.750000	0.000000	2.000000	19.000000	0.000000	0.000000	8.003
50%	78.500000	0.000000	3.000000	26.000000	0.000000	0.000000	14.454
75%	117.250000	1.000000	3.000000	35.000000	1.000000	0.000000	30.371
max	156.000000	1.000000	3.000000	71.000000	5.000000	5.000000	263.000

```
df.isna().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Lname            0
Name             0
Sex              0
Age             30
SibSp            0
Parch            0
```

```
Ticket           0
Fare             0
Cabin          125
Embarked         1
dtype: int64
```

```
df['Age'].fillna(df['Age'].mean(), inplace = True)
```

```
df['Cabin'].fillna(df['Cabin'].mode()[0], inplace = True)
```

```
df.isna().sum()
```

```
PassengerId    0
Survived       0
Pclass         0
Lname          0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       1
dtype: int64
```

Disk: 26.07 GB/107.72 GB

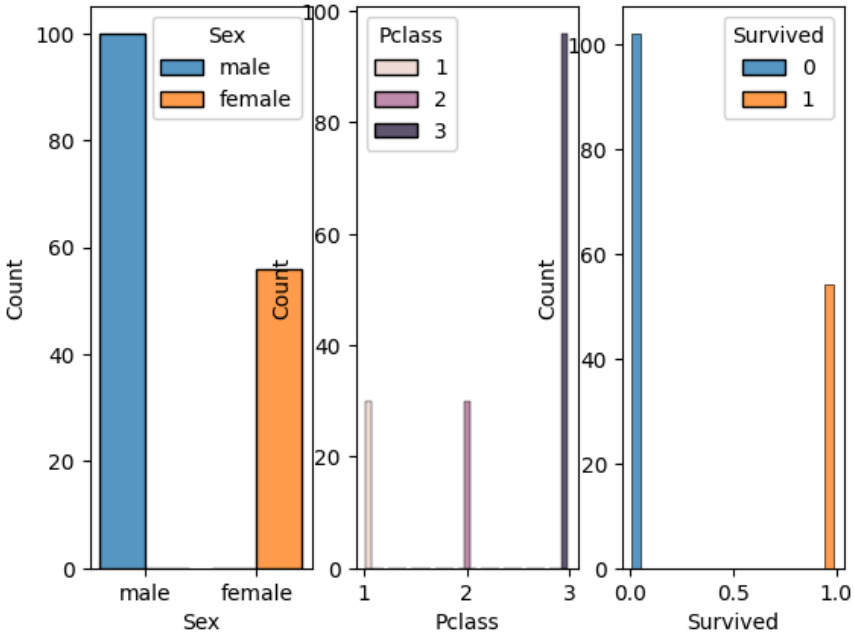
```
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace = True)
```

```
df.isna().sum()
```

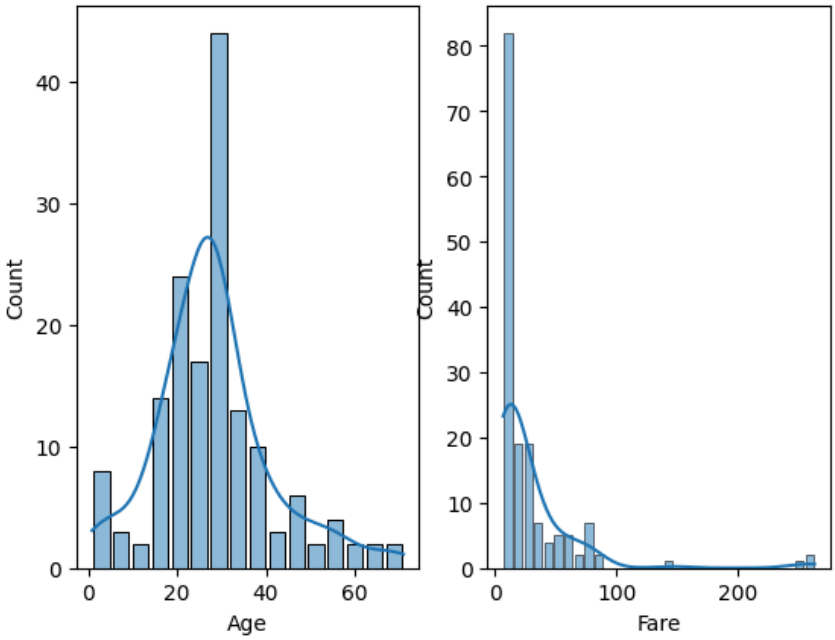
```
PassengerId    0
Survived       0
Pclass         0
Lname          0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

Single variable histogram

```
fig, axis = plt.subplots(1, 3)
sns.histplot(ax = axis[0], data = df, x='Sex', hue = 'Sex', multiple = 'dodge', shrink = 0.8)
sns.histplot(ax = axis[1], data = df, x='Pclass', hue = 'Pclass', multiple = 'dodge', shrink = 0.8)
sns.histplot(ax = axis[2], data = df, x='Survived', hue = 'Survived', multiple = 'dodge', shrink = 0.8)
plt.show()
```

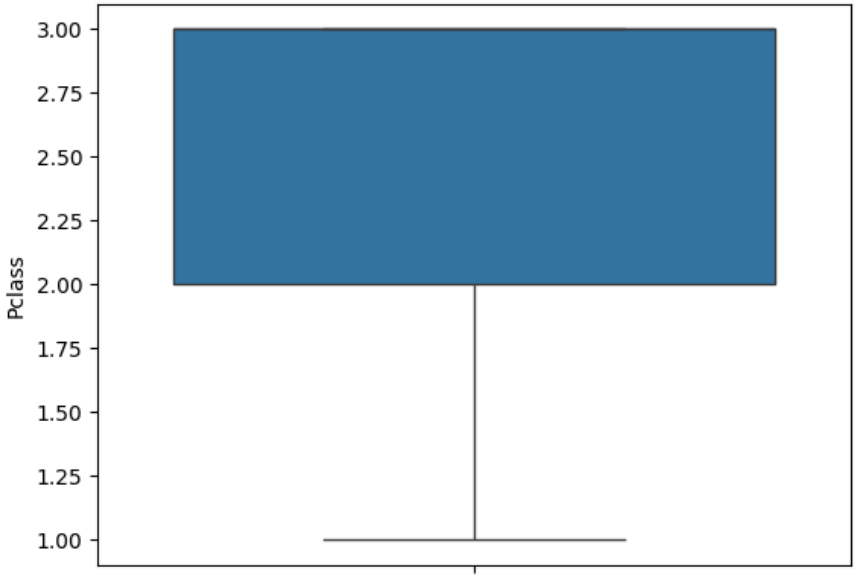



```
fig, axis = plt.subplots(1,2)
sns.histplot(ax = axis[0], data = df, x = 'Age', multiple = 'dodge', shrink = 0.8, kde = True)
sns.histplot(ax = ax                                Fare', multiple = 'dodge', shrink = 0.8, kde = True)
plt.show()
```

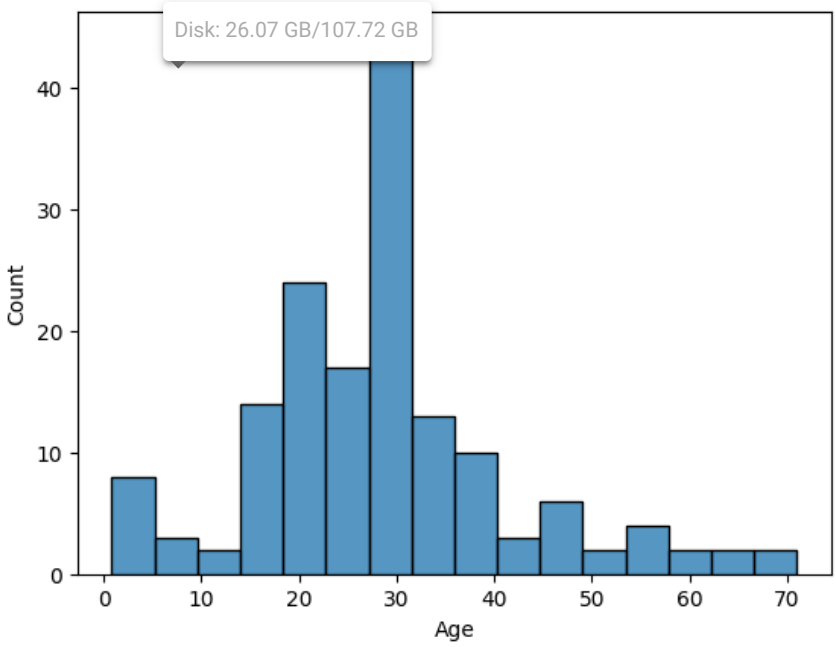

```
sns.boxplot(df['Pclass'])
```

```
<Axes: ylabel='Pclass'>
```



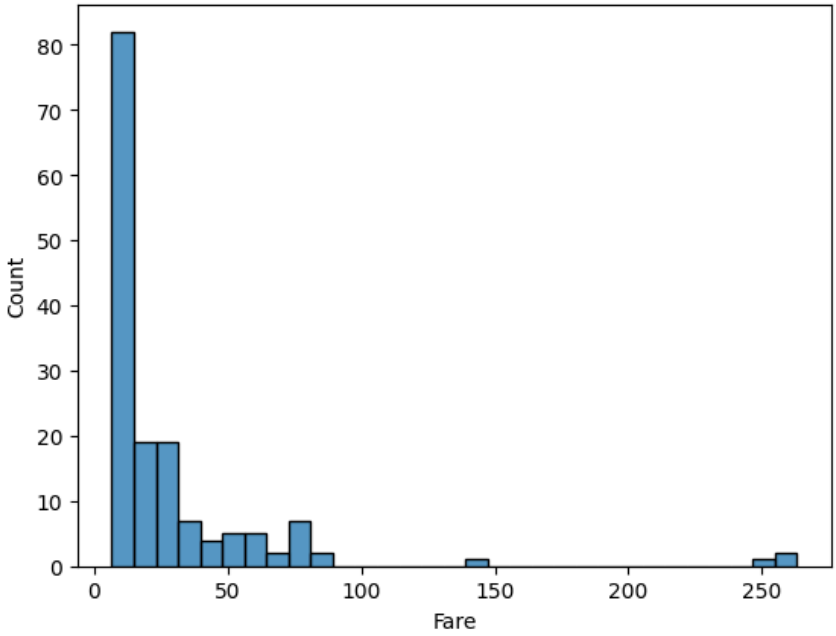

```
sns.histplot(df['Age'])
```

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




```
sns.histplot(df['Fare'])
```

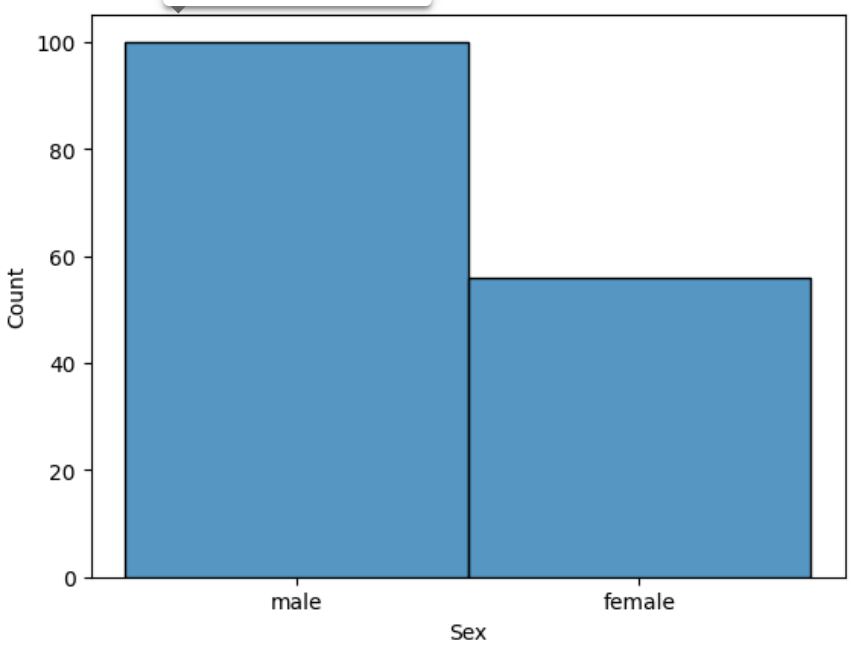
```
<Axes: xlabel='Fare', ylabel='Count'>
```




```
sns.histplot(df['Sex'])
```

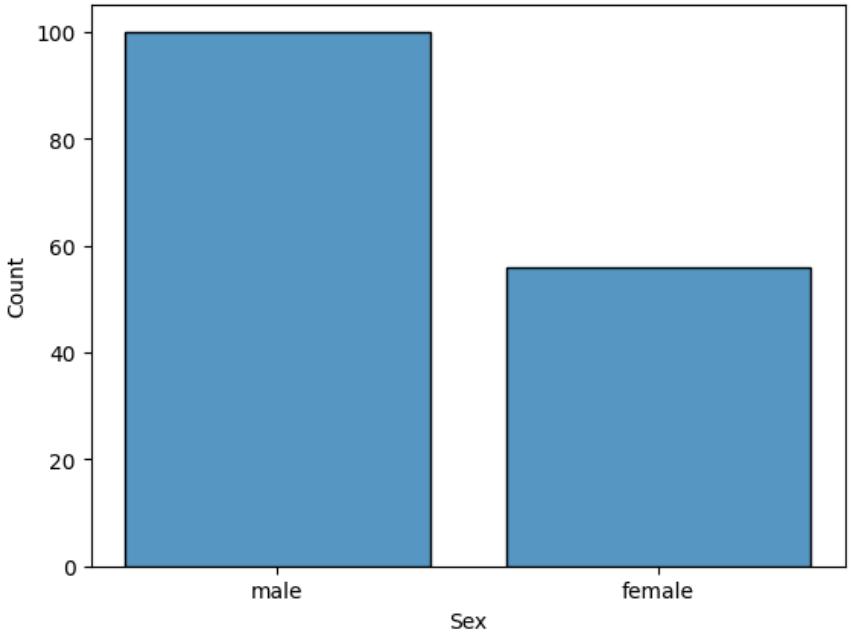
```
<Axes: xlabel='
```

Disk: 26.07 GB/107.72 GB




```
sns.histplot(df['Sex'], multiple = 'dodge', shrink = 0.8)
```

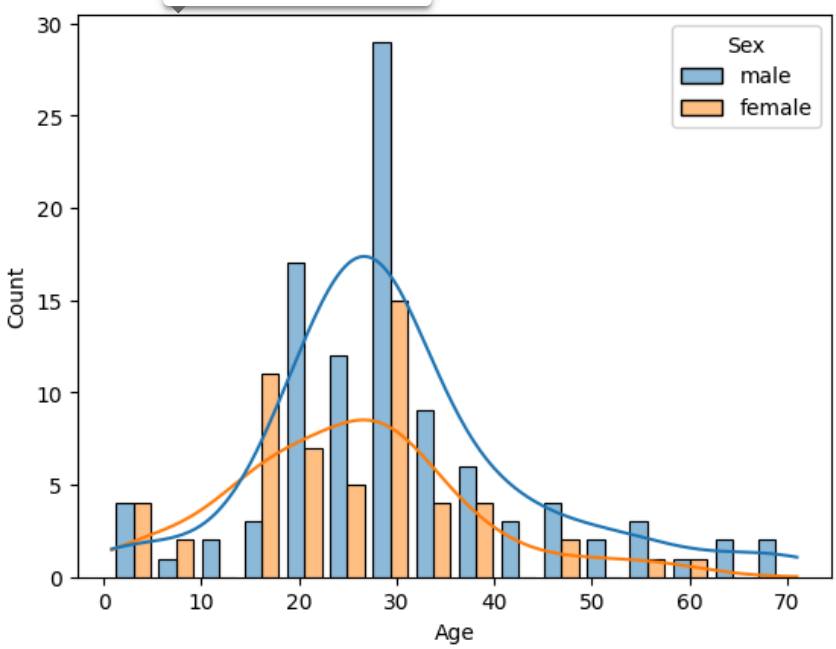
```
<Axes: xlabel='Sex', ylabel='Count'>
```




```
sns.histplot(data = df, x = 'Age', hue = 'Sex', multiple = 'dodge', shrink = 0.8, kde = True)
```

```
<Axes: xlabel='
```

Disk: 26.07 GB/107.72 GB

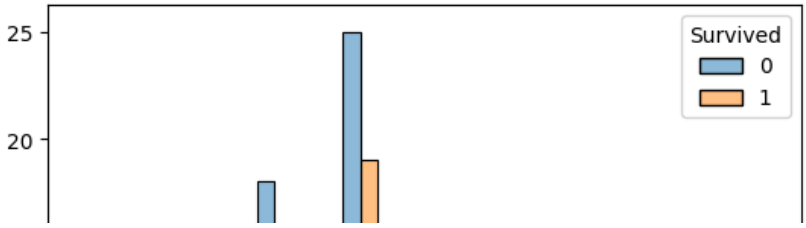



```
sns.histplot(data = df, x = 'Fare', hue = 'Sex', multiple = 'dodge', shrink = 0.8, kde = True)
```

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


```
sns.histplot(data = df, x = 'Age', hue = 'Survived', multiple = 'dodge', shrink = 0.8, kde = True)
```

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



Disk: 26.07 GB/107.72 GB