

. Data Visualization I

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

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

```
df=pd.read_csv('titanic.csv')
df
```

🔗

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket |
|---|-------------|----------|--------|--------------------------------------|--------|------|-------|-------|---------|
| 0 | 892 | 0 | 3 | Kelly, Mr. James | male | 34.5 | 0 | 0 | 330911 |
| 1 | 893 | 1 | 3 | Wilkes, Mrs. James (Ellen Needs) | female | 47.0 | 1 | 0 | 363272 |
| 2 | 894 | 0 | 2 | Myles, Mr. Thomas Francis | male | 62.0 | 0 | 0 | 240276 |
| 3 | 895 | 0 | 3 | Wirz, Mr. Albert | male | 27.0 | 0 | 0 | 315154 |
| 4 | 896 | 1 | 3 | Hirvonen, Mrs. Alexander (Helga E... | female | 22.0 | 1 | 1 | 3101298 |

Saved successfully!

✕

◀

▶

```
df.describe()
```

| | PassengerId | Survived | Pclass | Age | SibSp | Parch | Fare |
|-------|-------------|------------|------------|------------|------------|------------|------------|
| count | 418.000000 | 418.000000 | 418.000000 | 332.000000 | 418.000000 | 418.000000 | 417.000000 |
| mean | 1100.500000 | 0.363636 | 2.265550 | 30.272590 | 0.447368 | 0.392344 | 35.627188 |
| std | 120.810458 | 0.481622 | 0.841838 | 14.181209 | 0.896760 | 0.981429 | 55.907576 |
| min | 892.000000 | 0.000000 | 1.000000 | 0.170000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 996.250000 | 0.000000 | 1.000000 | 21.000000 | 0.000000 | 0.000000 | 7.895800 |
| 50% | 1100.500000 | 0.000000 | 3.000000 | 27.000000 | 0.000000 | 0.000000 | 14.454200 |
| 75% | 1204.750000 | 1.000000 | 3.000000 | 39.000000 | 1.000000 | 0.000000 | 31.500000 |
| max | 1309.000000 | 1.000000 | 3.000000 | 76.000000 | 8.000000 | 9.000000 | 512.329200 |

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  418 non-null    int64
1   Survived     418 non-null    int64
2   Pclass       418 non-null    int64
3   Name         418 non-null    object
4   Sex          418 non-null    object
5   Age          332 non-null    float64
6   SibSp        418 non-null    int64
7   Parch        418 non-null    int64
8   Ticket       418 non-null    object
9   Fare         417 non-null    float64
```

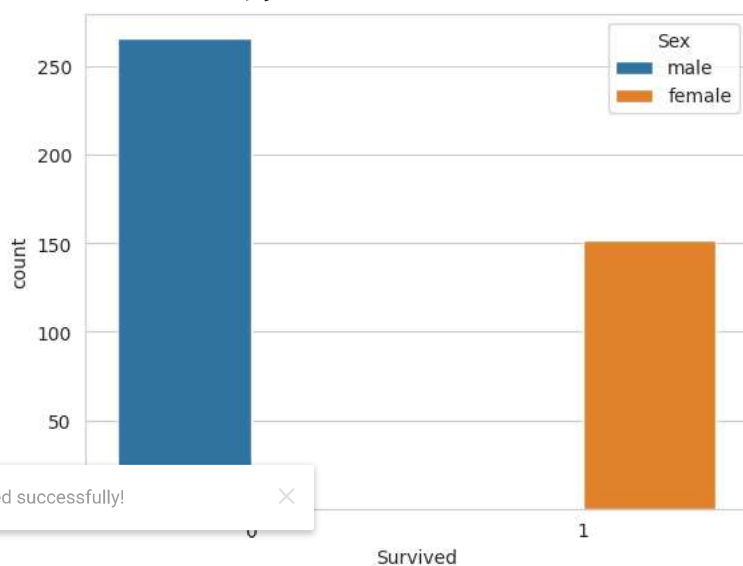
```
10 Cabin          91 non-null    object
11 Embarked       418 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

```
df.columns
```

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

```
sns.set_style('whitegrid')
sns.countplot(x='Survived', data=df, hue='Sex')
```

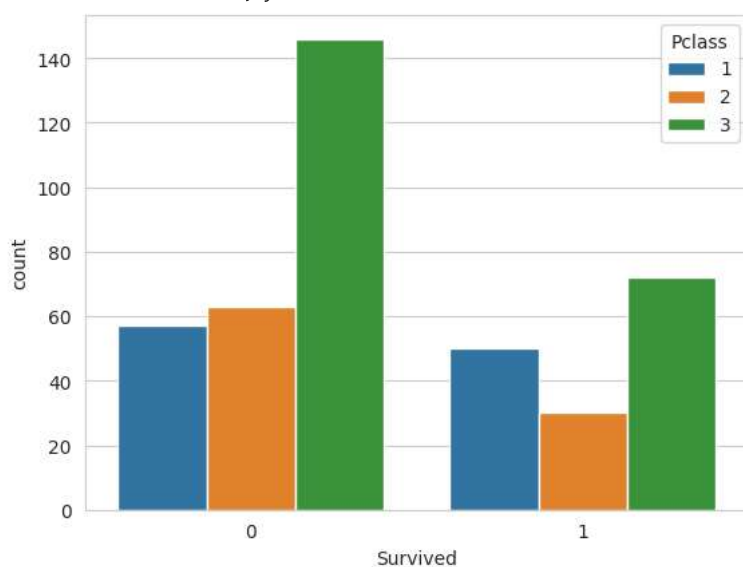
<Axes: xlabel='Survived', ylabel='count'>



Saved successfully!

```
sns.countplot(x='Survived', data=df, hue='Pclass')
```

<Axes: xlabel='Survived', ylabel='count'>



```
sns.distplot(df['Age'].dropna(), kde=False, bins=30)
```

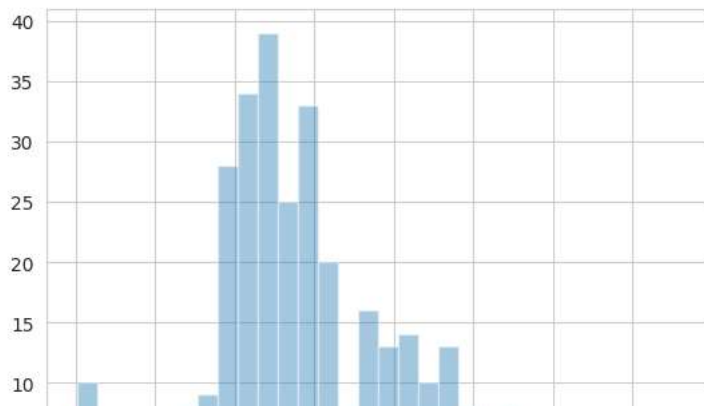
```
<ipython-input-10-554d3f2bc505>:1: UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

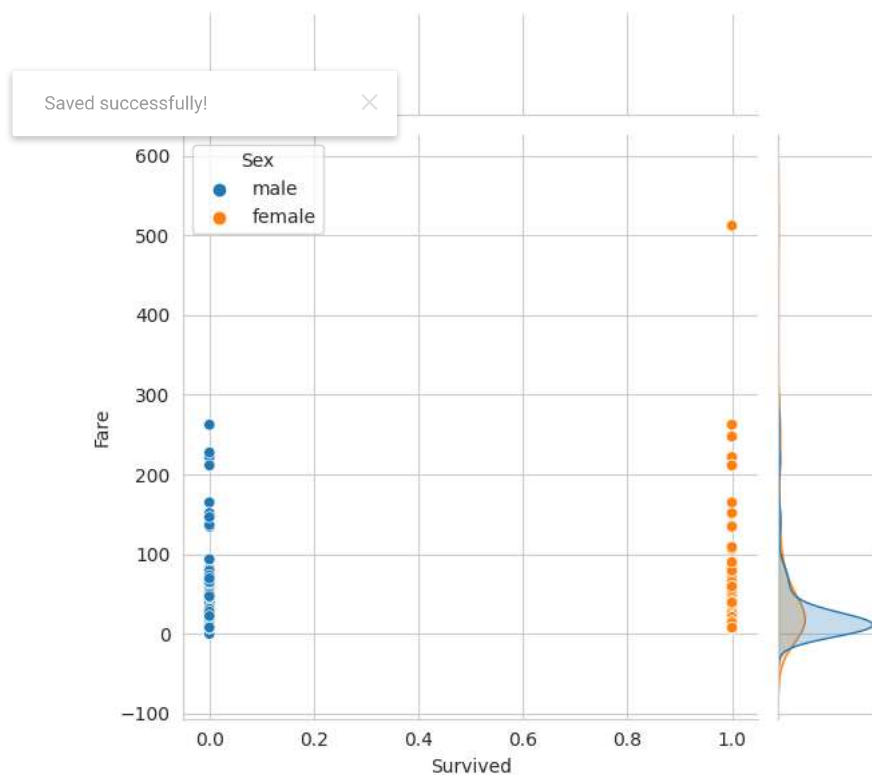
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['Age'].dropna(), kde=False, bins=30)
<Axes: xlabel='Age'>
```



```
sns.jointplot(x='Survived', y='Fare', data=df, hue='Sex')
```

```
<seaborn.axisgrid.JointGrid at 0x7f8aa37f7880>
```



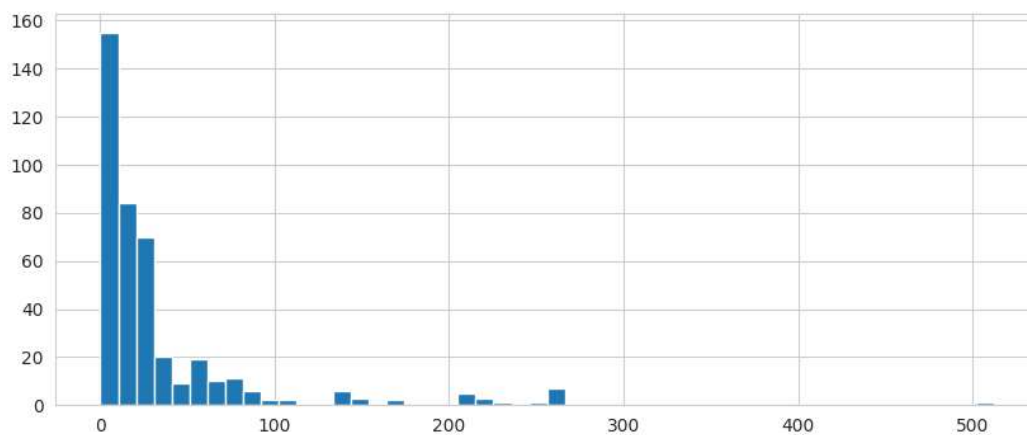
```
sns.countplot(x='SibSp', data=df)
```

<Axes: xlabel='SibSp', ylabel='count'>



```
df['Fare'].hist (bins=50, figsize=(10,4))
```

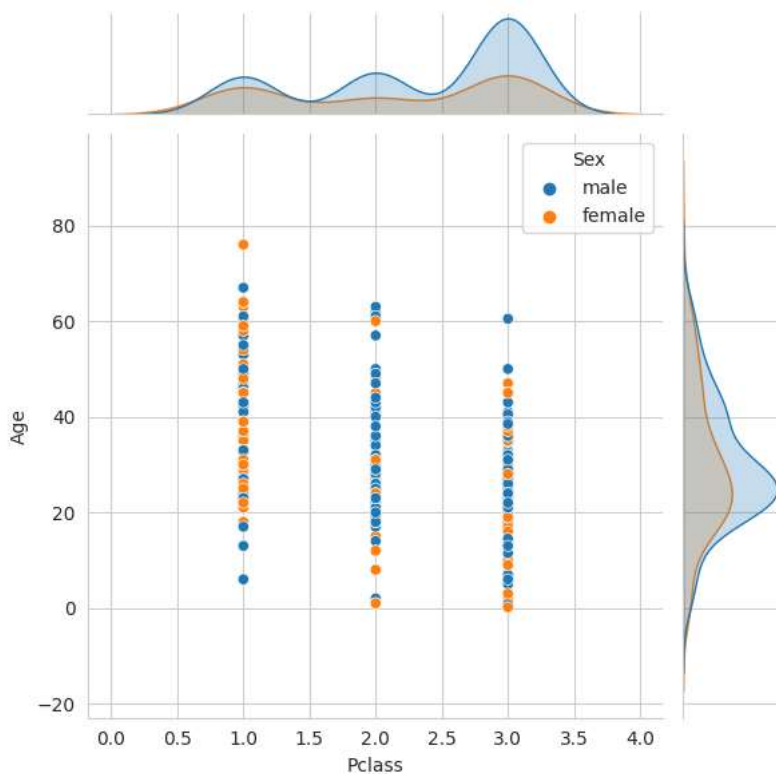
<Axes: >



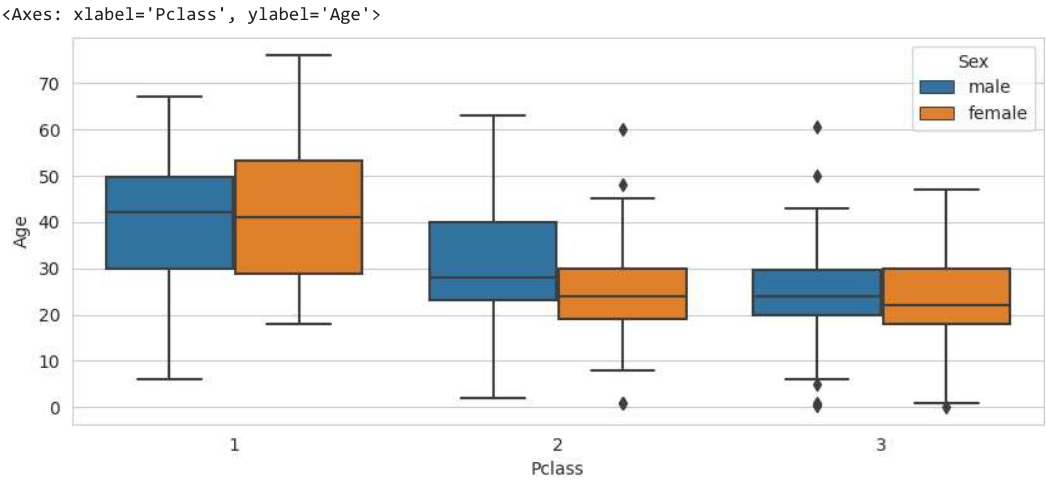
Saved successfully!

```
sns.jointplot(x='Pclass',y='Age', data=df, hue='Sex')
```

<seaborn.axisgrid.JointGrid at 0x7f8aa10d5630>



```
plt.figure(figsize=(10,4))
sns.boxplot(x='Pclass', y='Age', data=df, hue='Sex')
```



```
sns.pairplot(df)
```

Saved successfully! ×

```
df['Fare'].max()
```

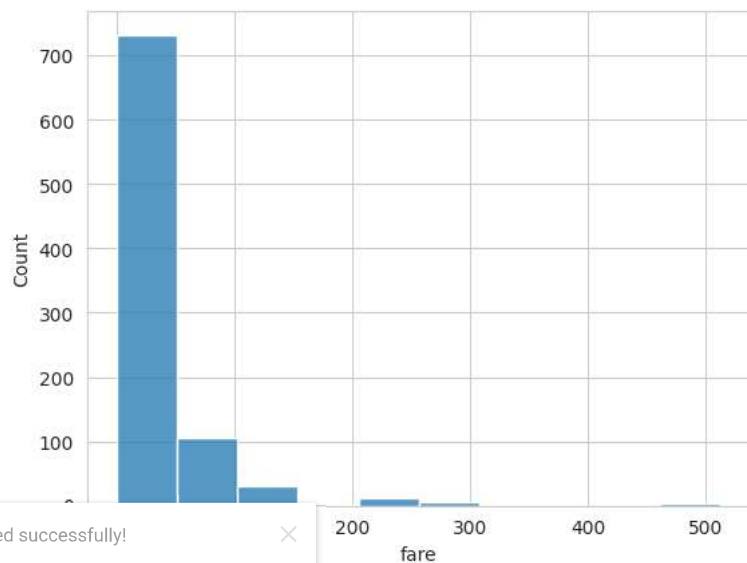
```
512.3292
```

```
df['Fare'].min()
```

```
0.0
```

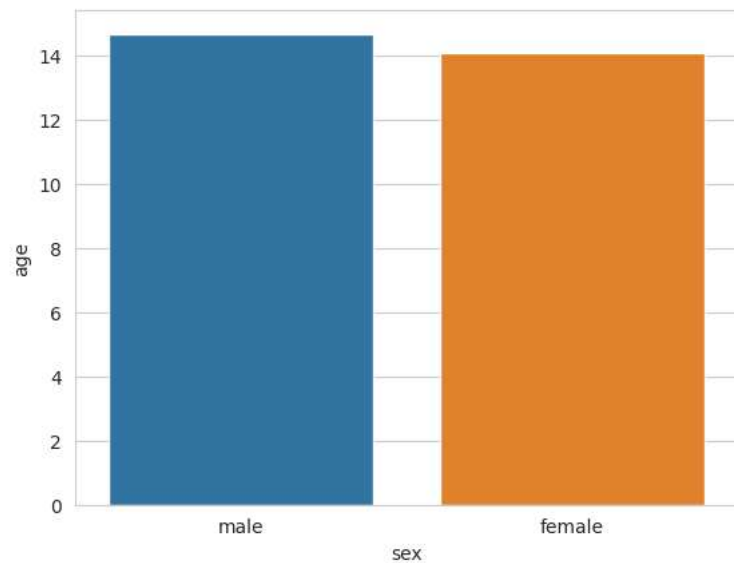
```
import seaborn as sns
df=sns.load_dataset('titanic')
sns.histplot(df['fare'],kde=False,bins=10)
```

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



```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.barplot(x='sex', y='age', data=df, estimator=np.std)
```

```
/usr/local/lib/python3.10/dist-packages/numpy/lib/nanfunctions.py:1560: RuntimeWarning: All-NaN slice encountered
r, k = function_base._ureduce(a,
/usr/local/lib/python3.10/dist-packages/numpy/lib/nanfunctions.py:1560: RuntimeWarning: All-NaN slice encountered
r, k = function_base._ureduce(a,
<Axes: xlabel='sex', ylabel='age'>
```





Saved successfully! 