```
Name : Suryal D . Khirade
Roll NO: T190424399
Assignment No :08
```

##Data Visualization I

1.Use the inbuilt dataset 'titanic'. The dataset contains 891 rows
and contains information about the passengers who boarded the unfo
rtunate Titanic ship. Use the Seaborn library to see if we can fin
d 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.

In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```python
titanic=sns.load_dataset("titanic")
```

In [3]:
```python
titanic
```

Out[3]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adul
0	0	3	male	22.0	1	0	7.2500	S	Third	man	
1	1	1	female	38.0	1	0	71.2833	C	First	woman	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	
3	1	1	female	35.0	1	0	53.1000	S	First	woman	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	
...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	
887	1	1	female	19.0	0	0	30.0000	S	First	woman	
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	
889	1	1	male	26.0	0	0	30.0000	C	First	man	
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	

891 rows × 15 columns

In [4]: `titanic.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```

In [5]: `titanic.shape`

Out[5]: (891, 15)

In [6]:
```python
titanic.describe
```

Out[6]:
```
<bound method NDFrame.describe of      survived  pclass    sex   age  si
bsp   parch     fare embarked  class  \
0                 0       3   male  22.0         1       0    7.2500       S   Thi
rd
1                 1       1 female  38.0         1       0   71.2833       C   Fir
st
2                 1       3 female  26.0         0       0    7.9250       S   Thi
rd
3                 1       1 female  35.0         1       0   53.1000       S   Fir
st
4                 0       3   male  35.0         0       0    8.0500       S   Thi
rd
..              ...     ...    ...   ...       ...     ...       ...     ...
...
886               0       2   male  27.0         0       0   13.0000       S   Seco
nd
887               1       1 female  19.0         0       0   30.0000       S   Fir
st
888               0       3 female   NaN         1       2   23.4500       S   Thi
rd
889               1       1   male  26.0         0       0   30.0000       C   Fir
st
890               0       3   male  32.0         0       0    7.7500       Q   Thi
rd

         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
..       ...         ...  ...          ...   ...    ...
886      man        True  NaN  Southampton    no   True
887    woman       False    B  Southampton   yes   True
888    woman       False  NaN  Southampton    no  False
889      man        True    C    Cherbourg   yes   True
890      man        True  NaN   Queenstown    no   True

[891 rows x 15 columns]>
```

In [7]:
```
titanic.describe
```

Out[7]:
```
<bound method NDFrame.describe of      survived   pclass    sex   age   si
bsp   parch     fare  embarked    class \
0          0        3   male  22.0      1      0   7.2500         S   Thi
rd
1          1        1 female  38.0      1      0  71.2833         C   Fir
st
2          1        3 female  26.0      0      0   7.9250         S   Thi
rd
3          1        1 female  35.0      1      0  53.1000         S   Fir
st
4          0        3   male  35.0      0      0   8.0500         S   Thi
rd
..       ...      ...    ...   ...    ...    ...      ...       ...
...
886        0        2   male  27.0      0      0  13.0000         S  Seco
nd
887        1        1 female  19.0      0      0  30.0000         S   Fir
st
888        0        3 female   NaN      1      2  23.4500         S   Thi
rd
889        1        1   male  26.0      0      0  30.0000         C   Fir
st
890        0        3   male  32.0      0      0   7.7500         Q   Thi
rd

        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
..      ...         ...   ...           ...   ...    ...
886     man        True   NaN   Southampton    no   True
887   woman       False     B   Southampton   yes   True
888   woman       False   NaN   Southampton    no  False
889     man        True     C     Cherbourg   yes   True
890     man        True   NaN    Queenstown    no   True

[891 rows x 15 columns]>
```

In [8]:
```python
titanic.isnull().sum()
```

Out[8]:
```
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
```
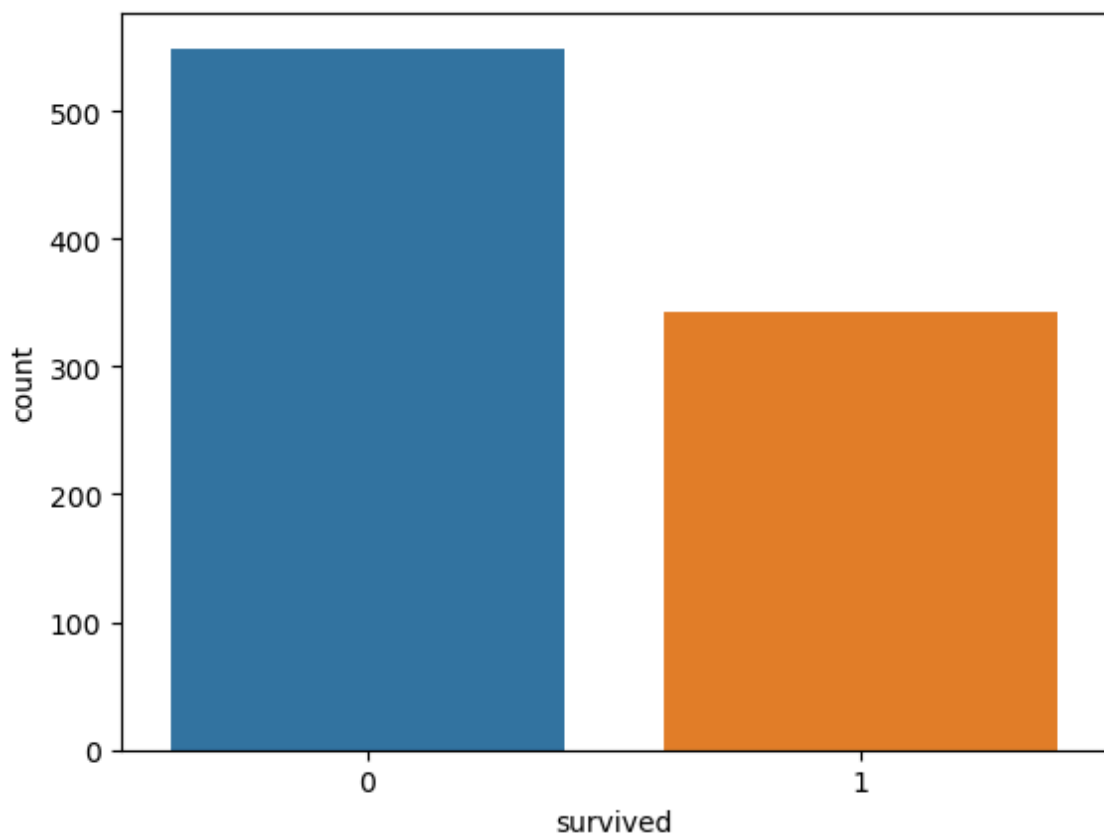
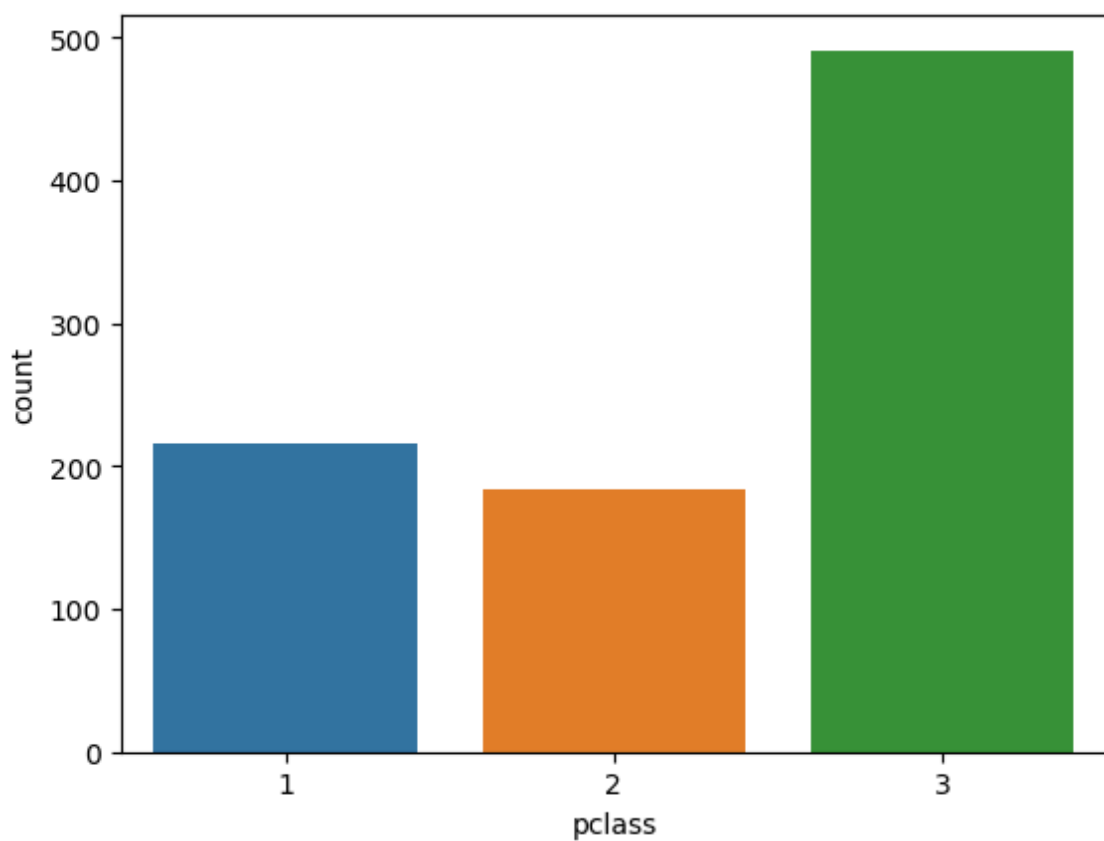In [9]:
```python
sns.countplot(x ='sex', data = titanic)
```

Out[9]: `<Axes: xlabel='sex', ylabel='count'>`
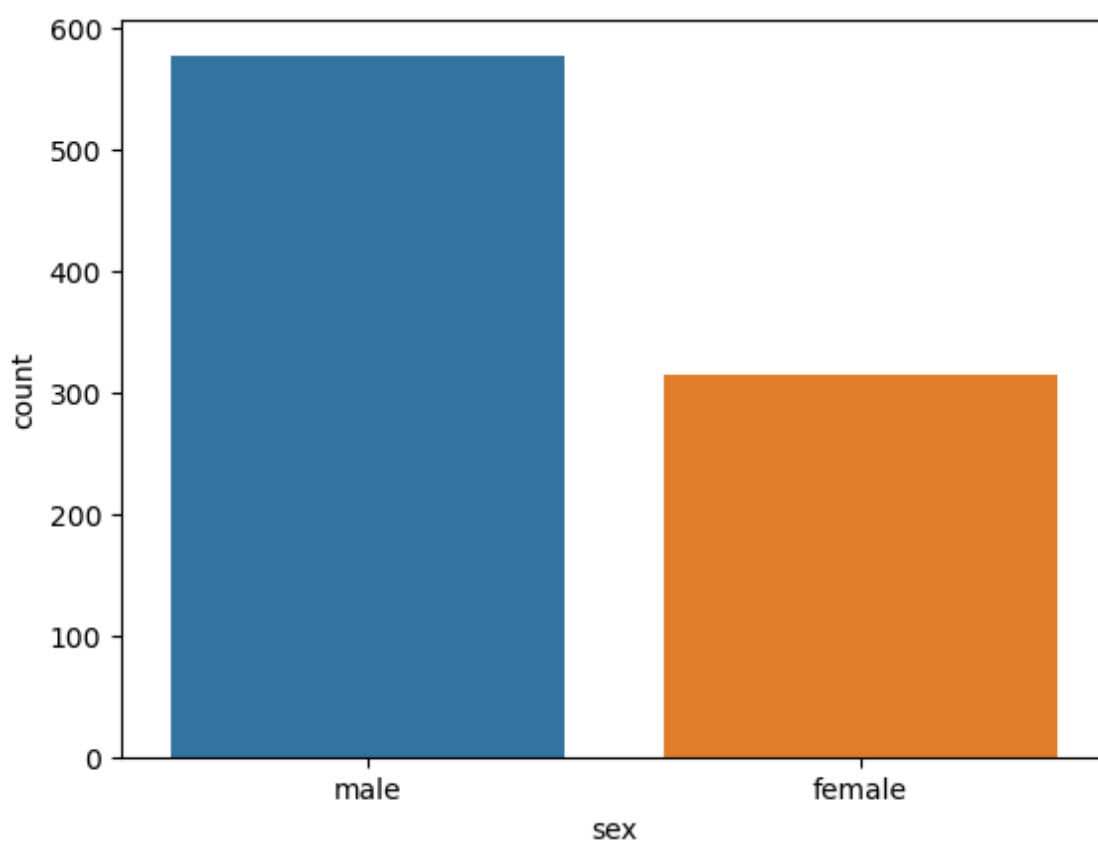
In [10]: `sns.countplot(x ='survived', data =titanic)`

Out[10]: `<Axes: xlabel='survived', ylabel='count'>`

In [11]: `sns.countplot(x ='pclass', data =titanic)`

Out[11]: `<Axes: xlabel='pclass', ylabel='count'>`

In [12]: `sns.countplot(x ='embark_town', data =titanic)`

Out[12]: `<Axes: xlabel='embark_town', ylabel='count'>`
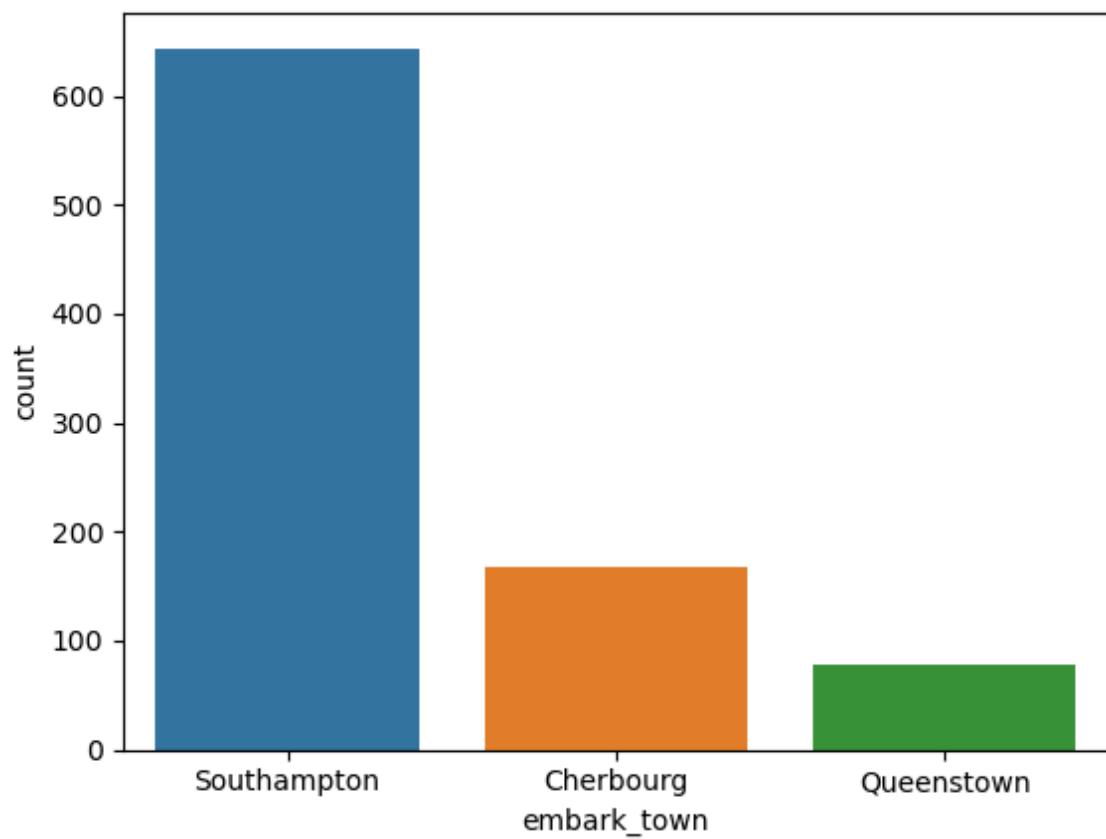
In [13]: `sns.boxplot(titanic.age)`

Out[13]: `<Axes: >`
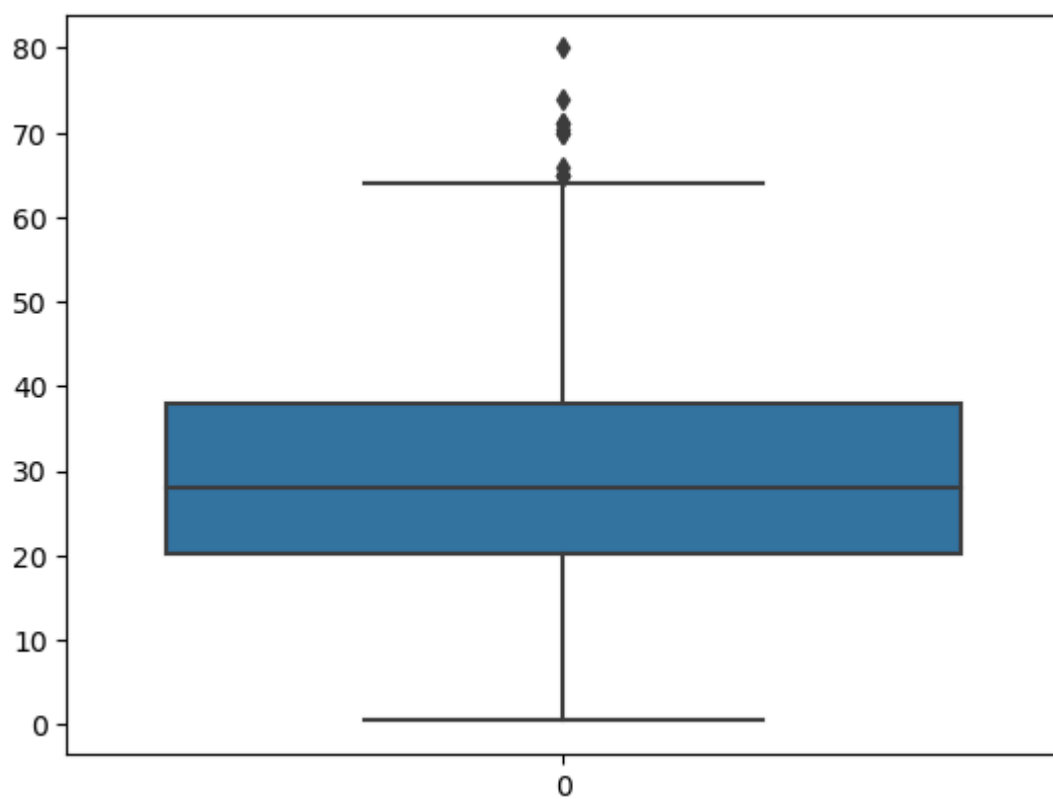
In [14]: `sns.boxplot(titanic.fare)`

Out[14]: `<Axes: >`
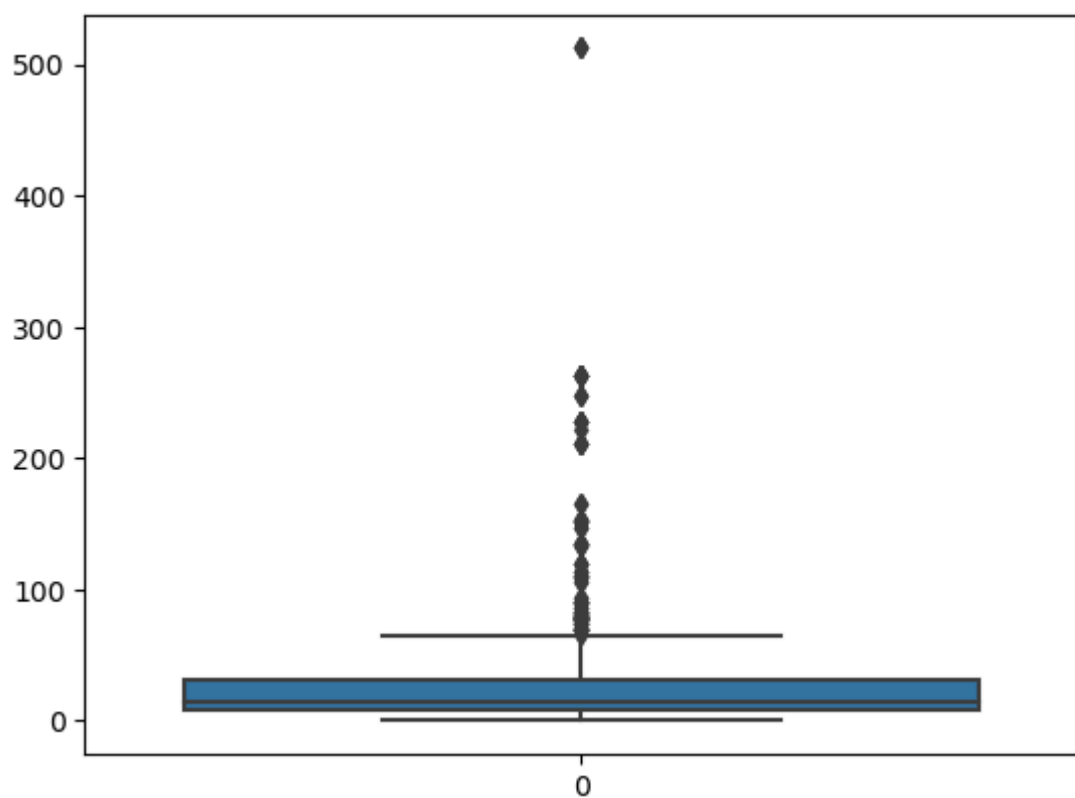
In [15]: `sns.boxplot(titanic.pclass)`

Out[15]: `<Axes: >`
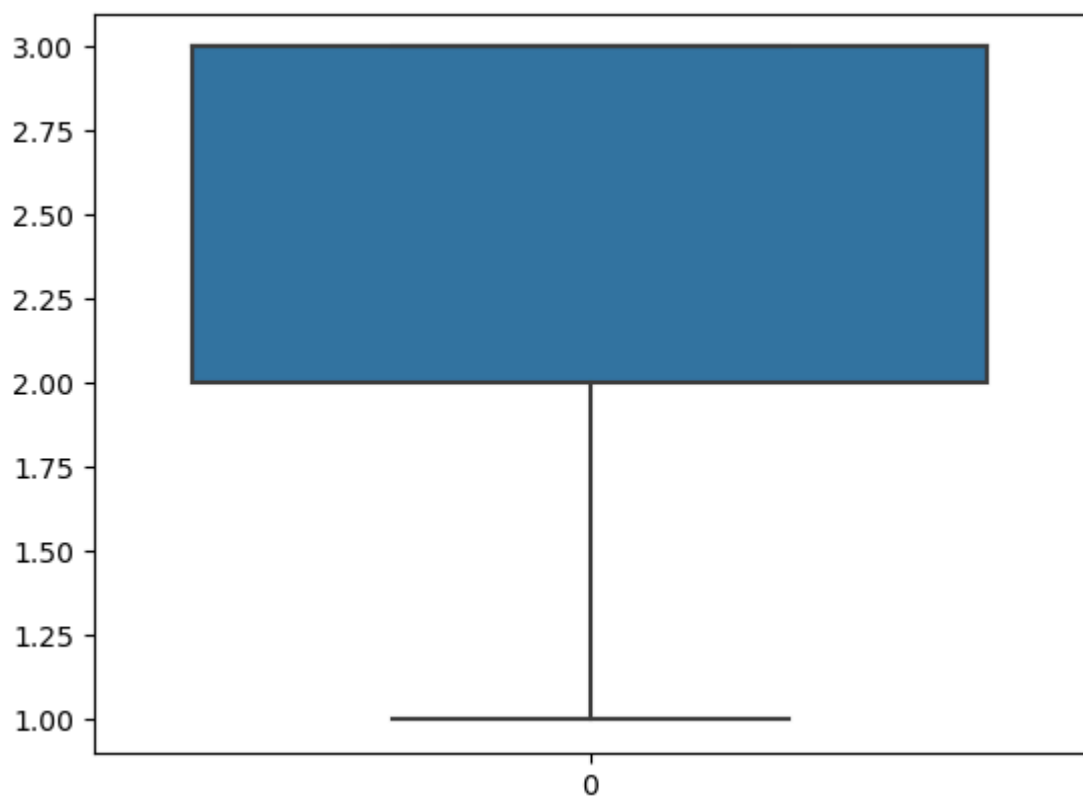
In [16]: `sns.boxplot(data = titanic , x= "pclass" ,y = "age")`

Out[16]: `<Axes: xlabel='pclass', ylabel='age'>`
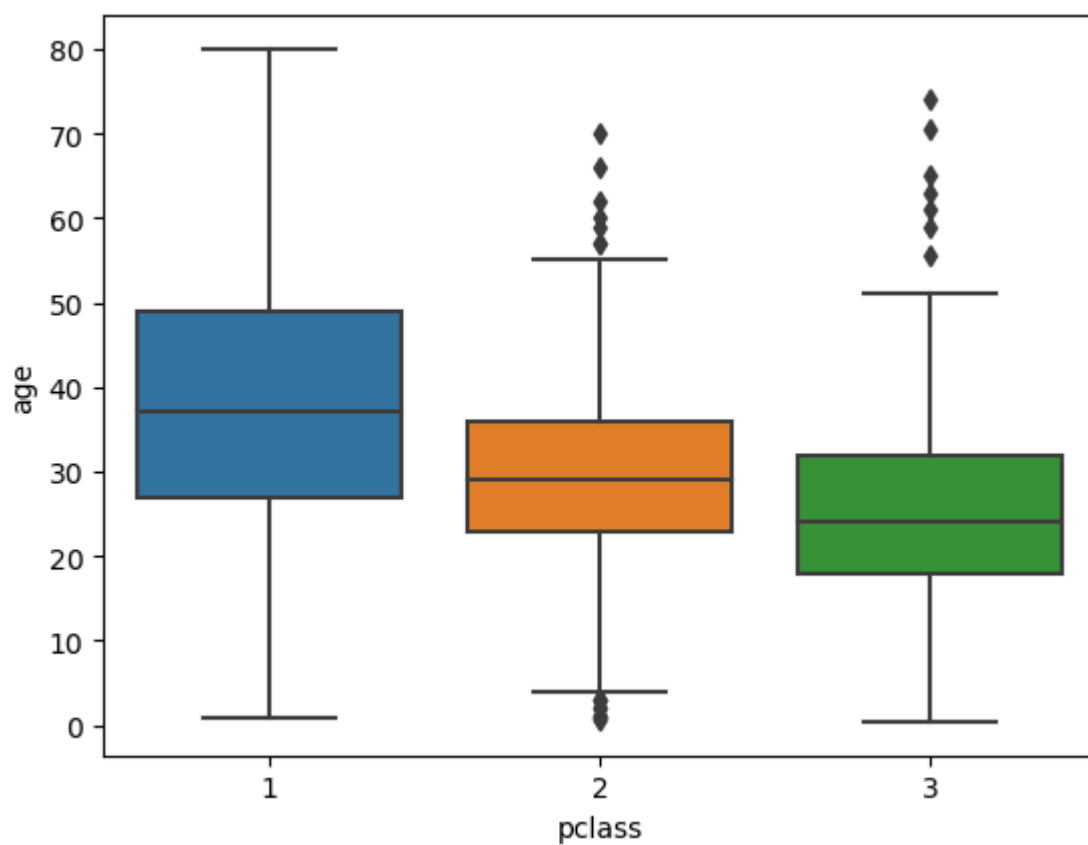
In [17]: `sns.catplot(x='pclass', y='fare', data=titanic)`

```
C:\Users\alisu\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserW
arning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[17]: `<seaborn.axisgrid.FacetGrid at 0x200f93279d0>`
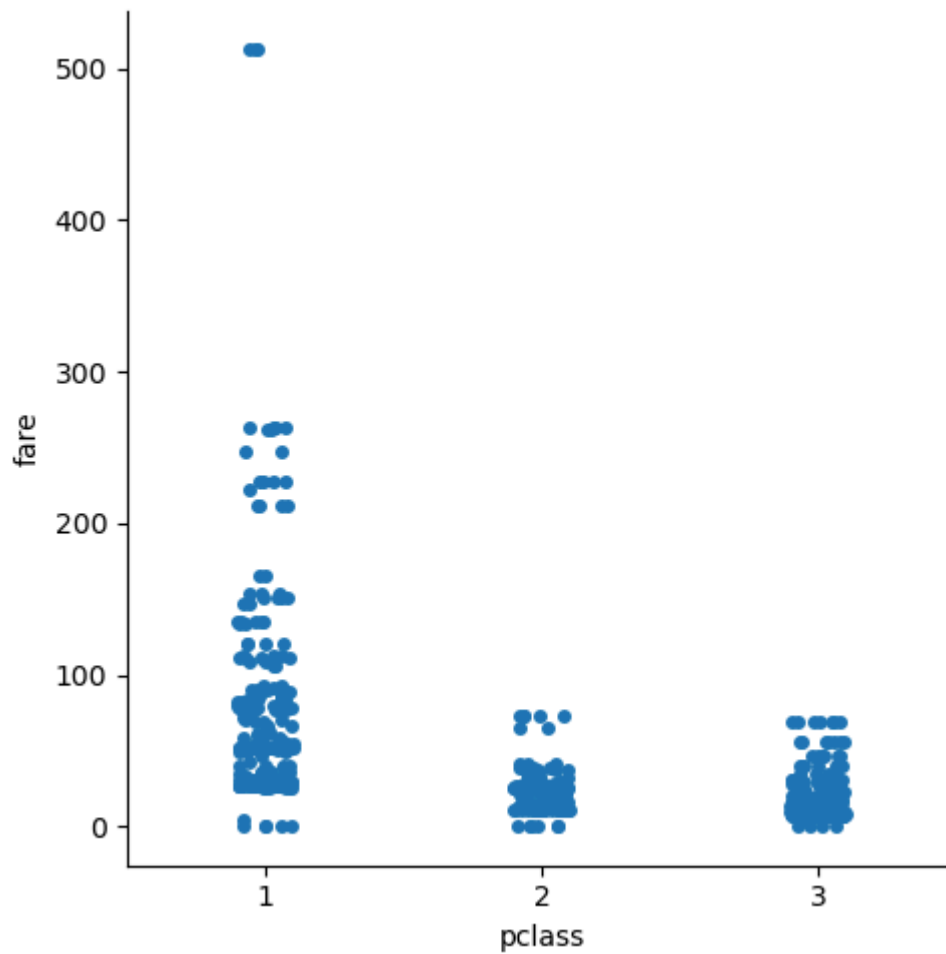
In [18]: `sns.catplot(x='sex', y='fare', data=titanic)`

```
C:\Users\alisu\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserW
arning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[18]: `<seaborn.axisgrid.FacetGrid at 0x200f8a02f90>`
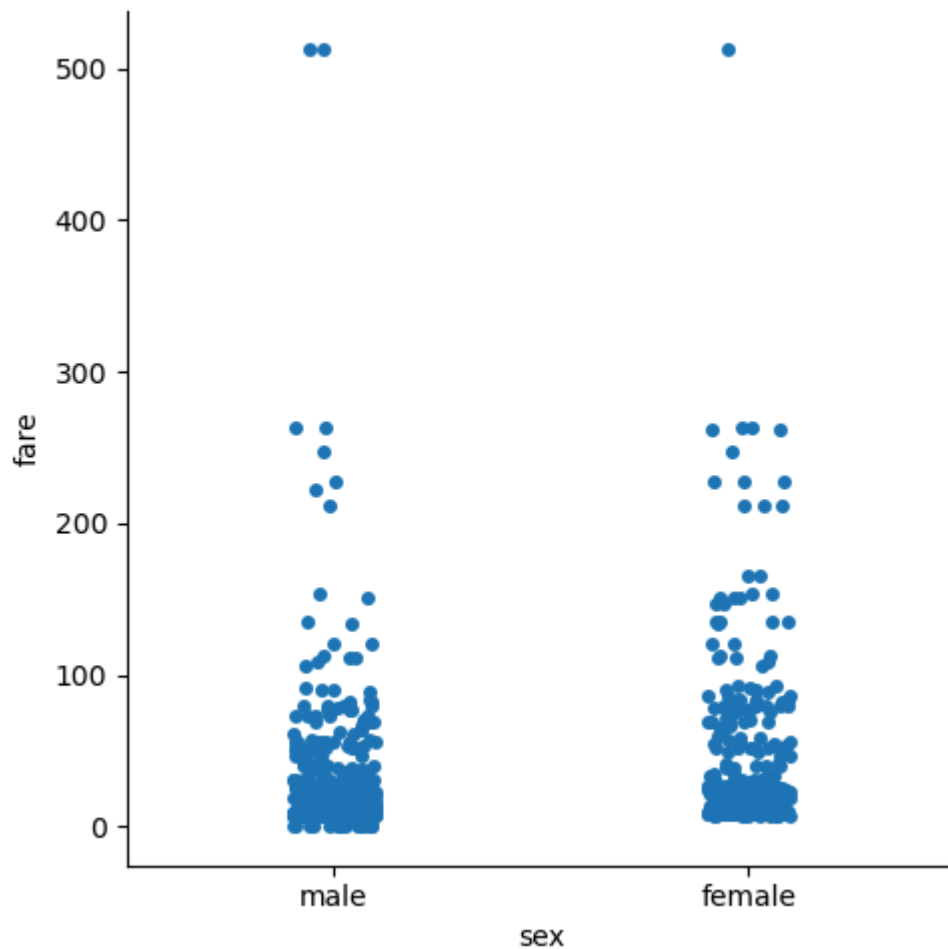
In [19]: `sns.catplot(x='age', y='sex', data=titanic)`

```
C:\Users\alisu\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserW
arning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[19]: `<seaborn.axisgrid.FacetGrid at 0x200f9285050>`
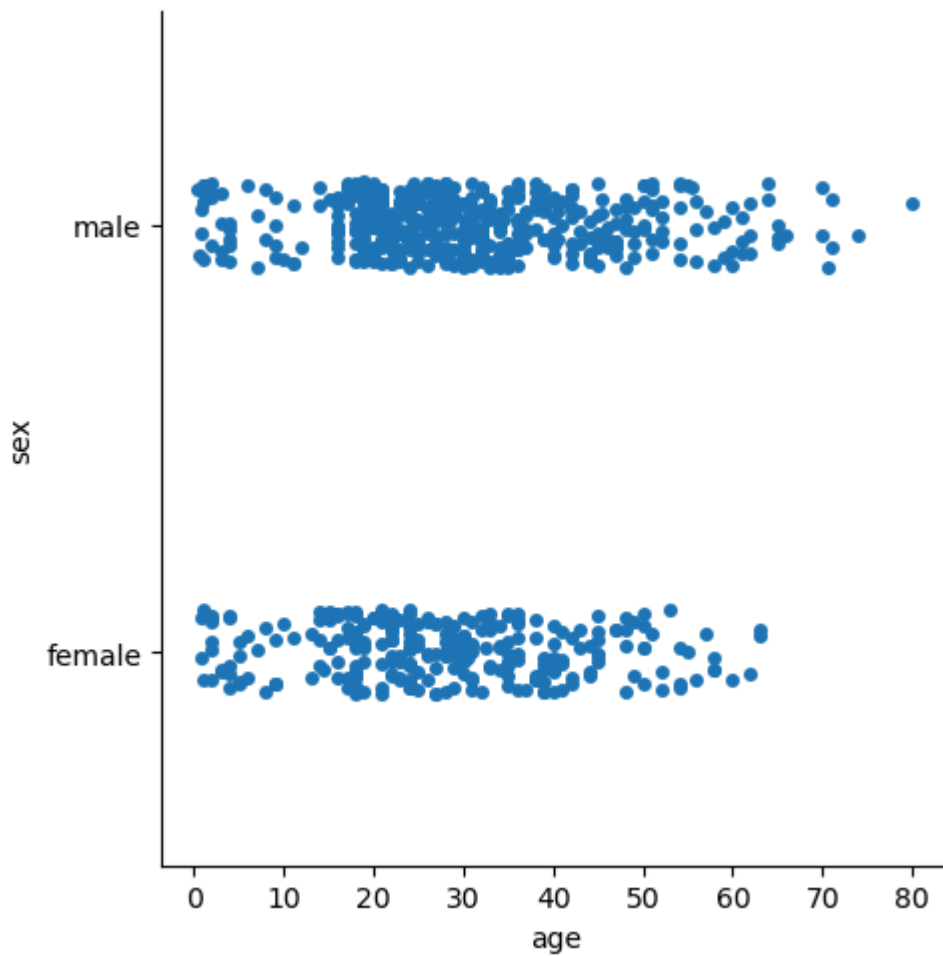
In [20]: `sns.pairplot(titanic)`

```
<__array_function__ internals>:200: RuntimeWarning: Converting input from
bool to <class 'numpy.uint8'> for compatibility.
<__array_function__ internals>:200: RuntimeWarning: Converting input from
bool to <class 'numpy.uint8'> for compatibility.
C:\Users\alisu\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserW
arning: The figure layout has changed to tight
    self._figure.tight_layout(*args, **kwargs)
```

Out[20]: `<seaborn.axisgrid.PairGrid at 0x200fa5a1250>`
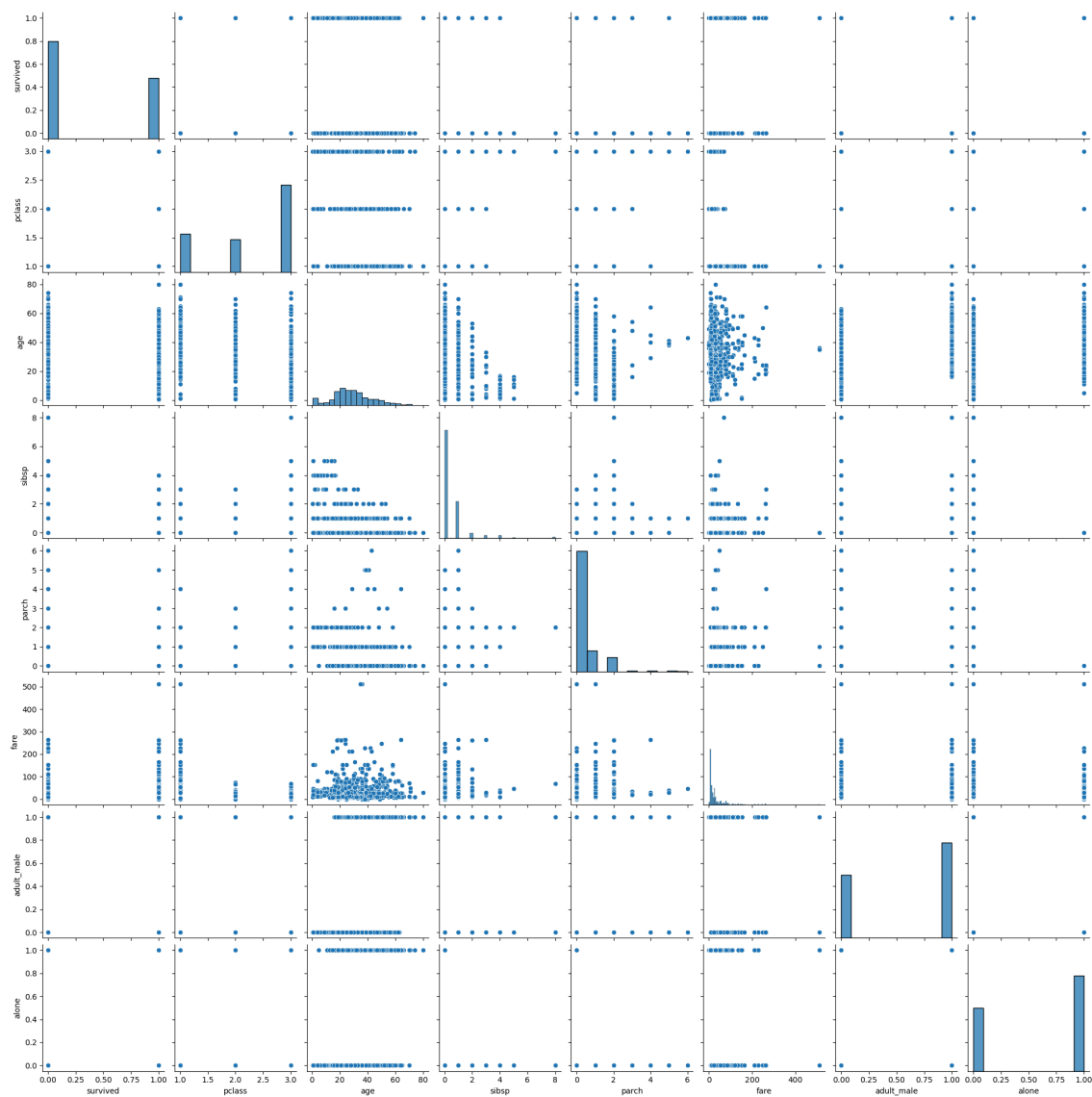
In [21]: `sns.scatterplot(x='fare',y='pclass', data = titanic,hue= 'survived')`

Out[21]: `<Axes: xlabel='fare', ylabel='pclass'>`
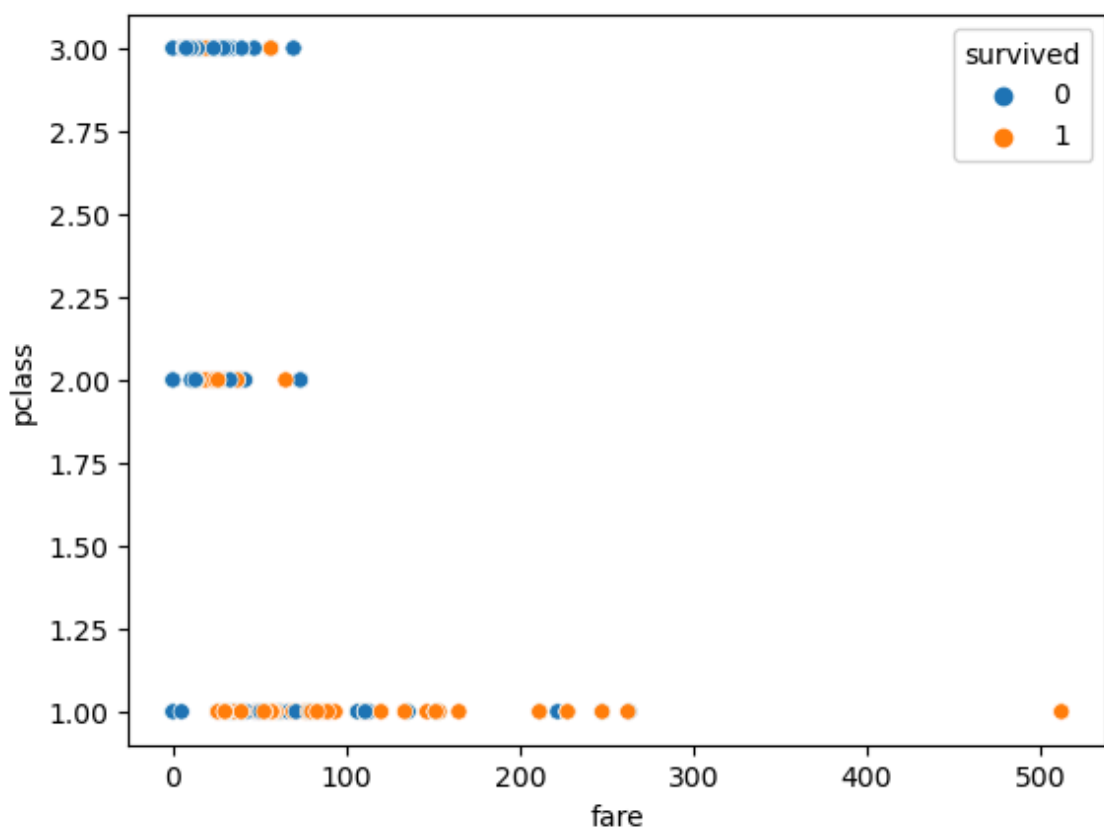
In [22]: `sns.distplot(titanic.fare)`

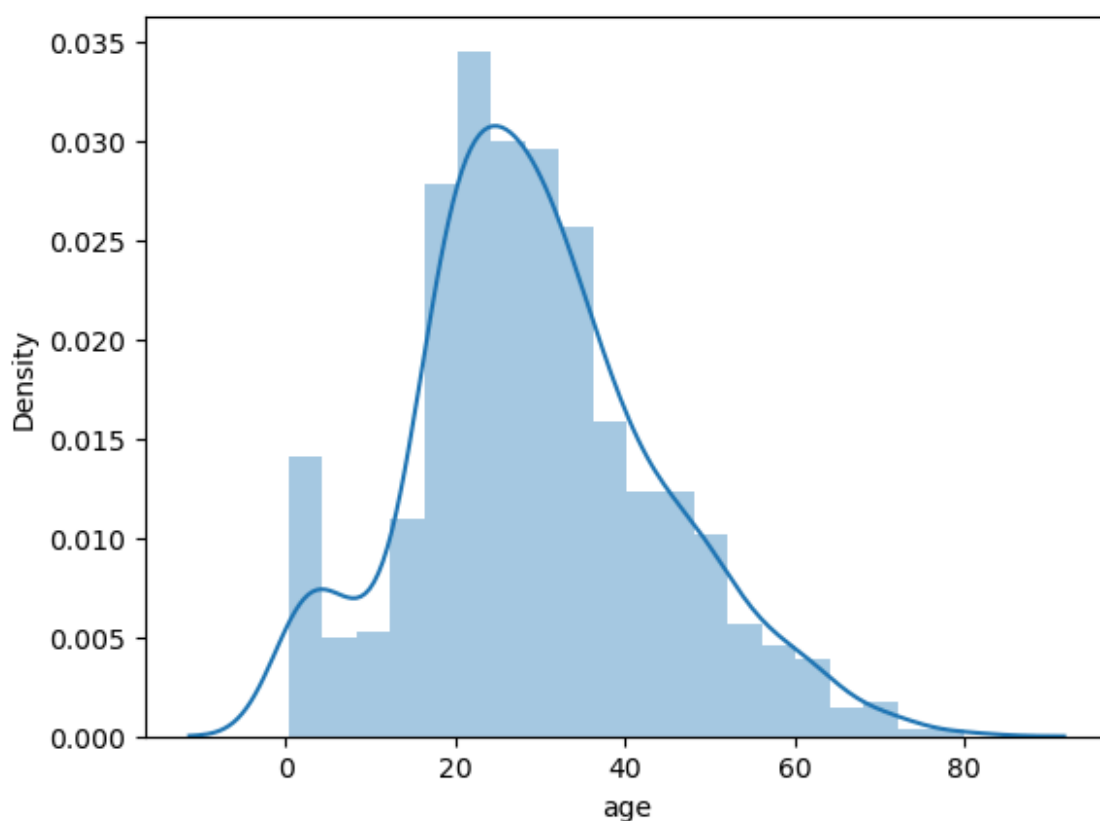C:\Users\alisu\AppData\Local\Temp\ipykernel_10124\3635634896.py:1: UserWa
rning:

`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 w
ith
similar flexibility) or `histplot` (an axes-level function for histogram
s).

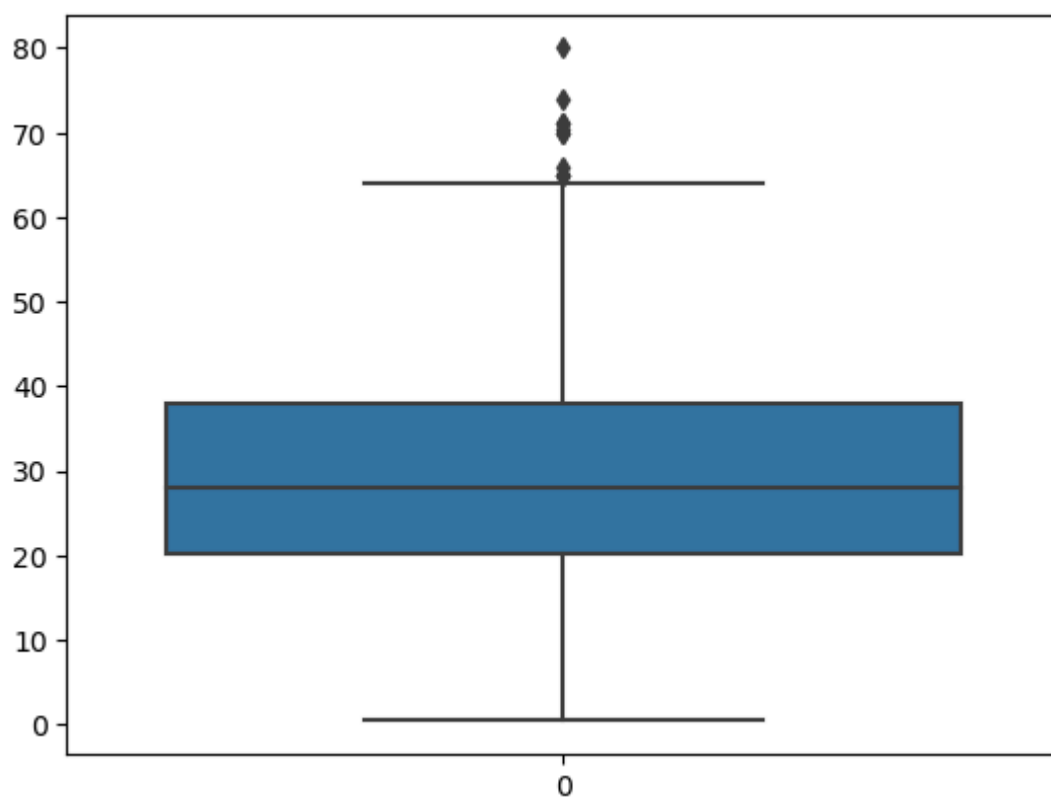For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (http
s://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot(titanic.fare)

Out[22]: `<Axes: xlabel='fare', ylabel='Density'>`

In [23]: `sns.distplot(titanic.age)`

C:\Users\alisu\AppData\Local\Temp\ipykernel_10124\3497382289.py:1: UserWa
rning:

`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 w
ith
similar flexibility) or `histplot` (an axes-level function for histogram
s).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (http
s://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot(titanic.age)

Out[23]: <Axes: xlabel='age', ylabel='Density'>

In [24]: `sns.boxplot(titanic.age)`
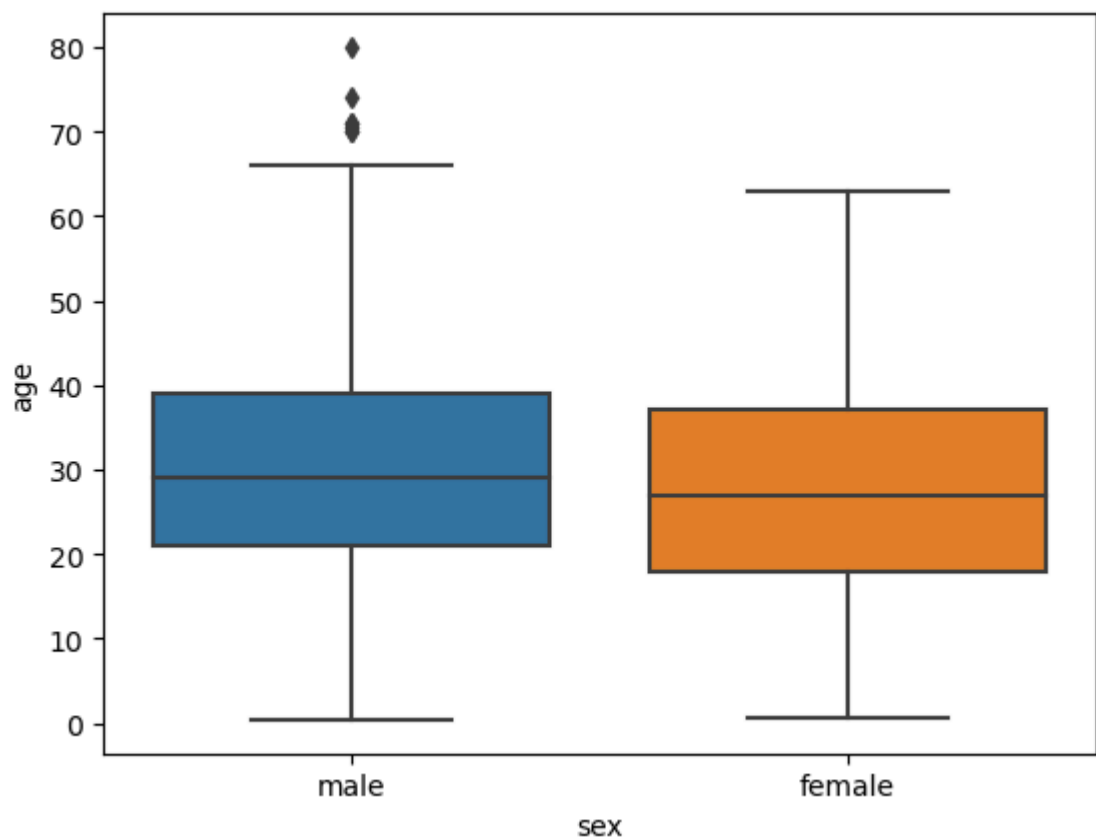
Out[24]: `<Axes: >`

In [25]: `sns.boxplot(titanic.fare)`
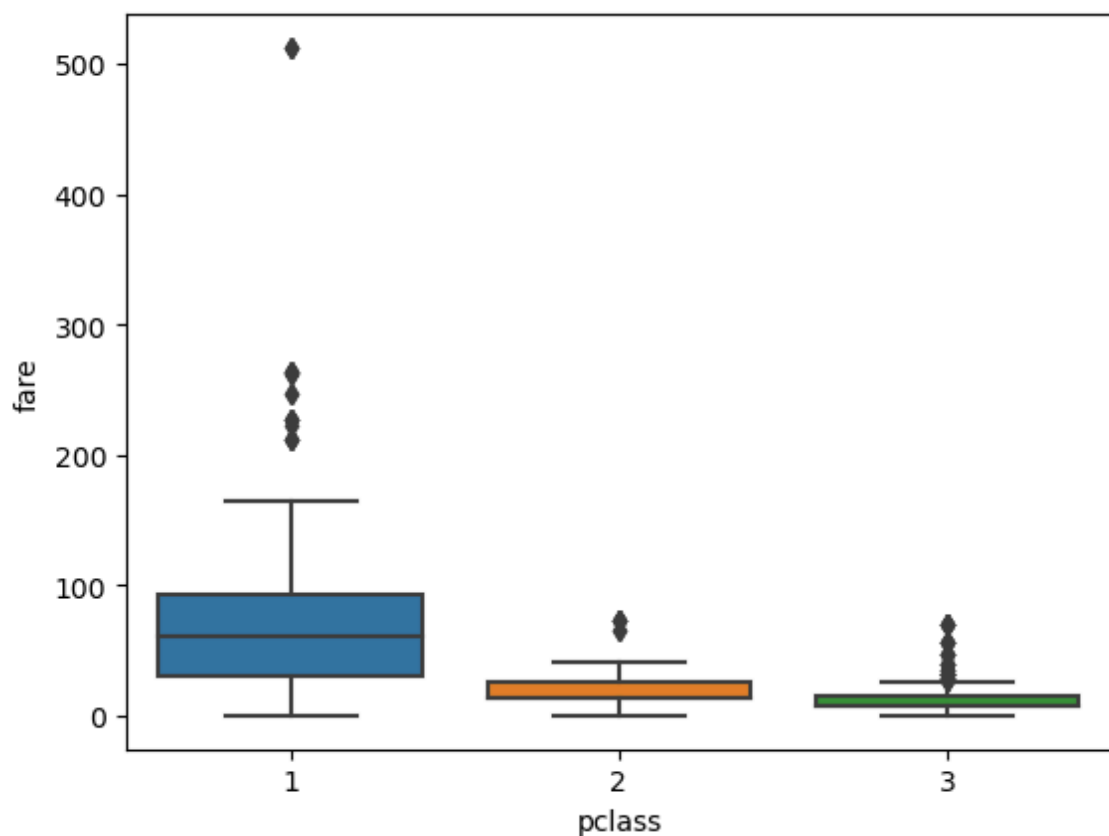
Out[25]: `<Axes: >`
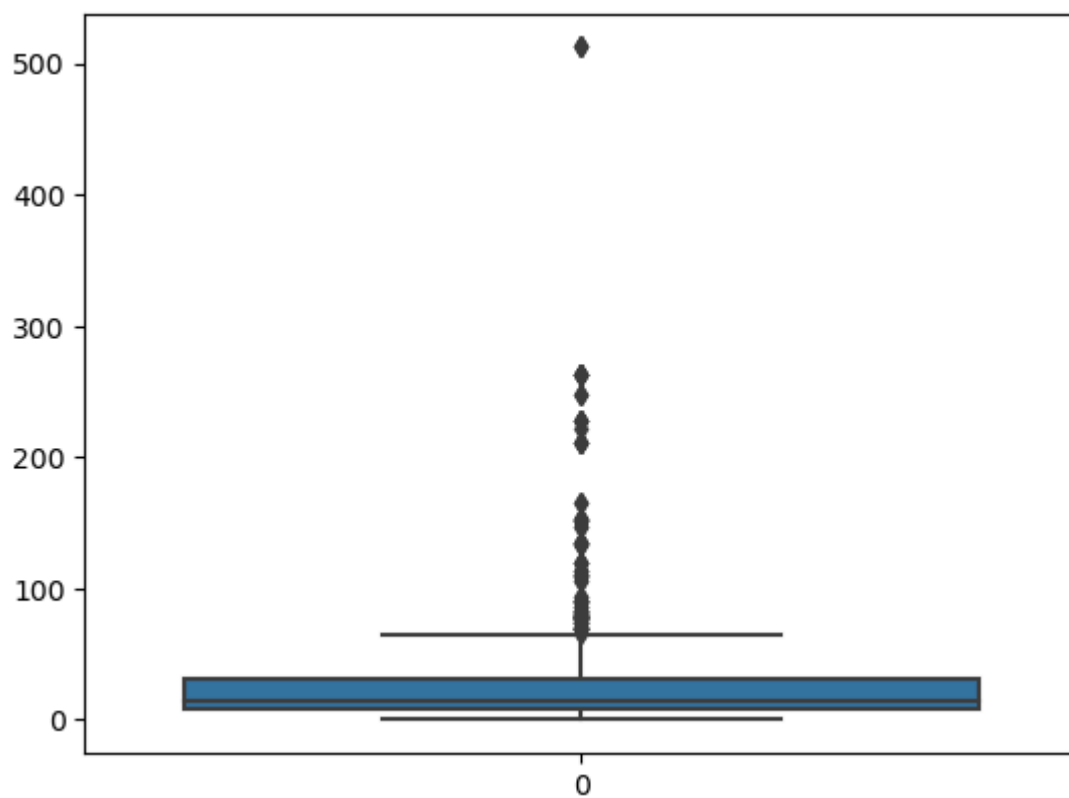
In [26]: `sns.boxplot(data = titanic , x= "sex" ,y = "age" )`

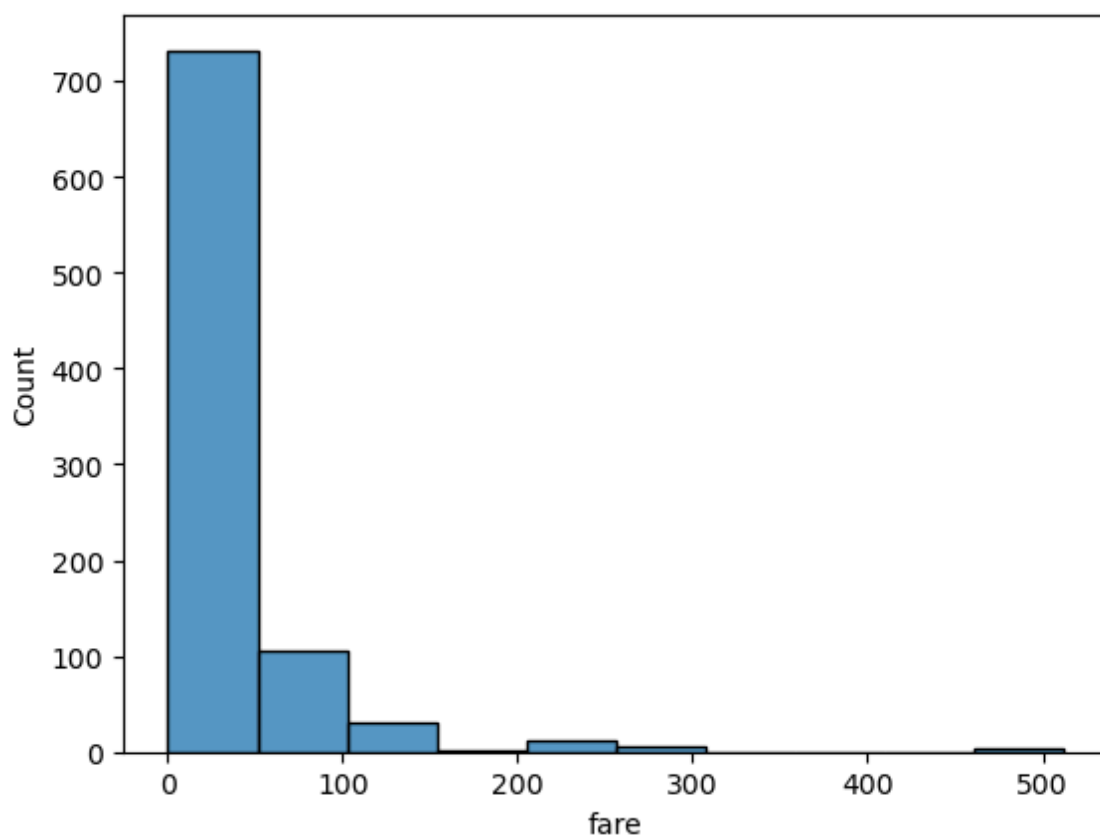Out[26]: `<Axes: xlabel='sex', ylabel='age'>`

In [27]: `sns.boxplot(data = titanic , x= "pclass" ,y = "fare" )`

Out[27]: `<Axes: xlabel='pclass', ylabel='fare'>`

In [28]: `sns.histplot(titanic.fare ,bins=10)`

Out[28]: `<Axes: xlabel='fare', ylabel='Count'>`

In [29]: `sns.distplot(titanic.age)`

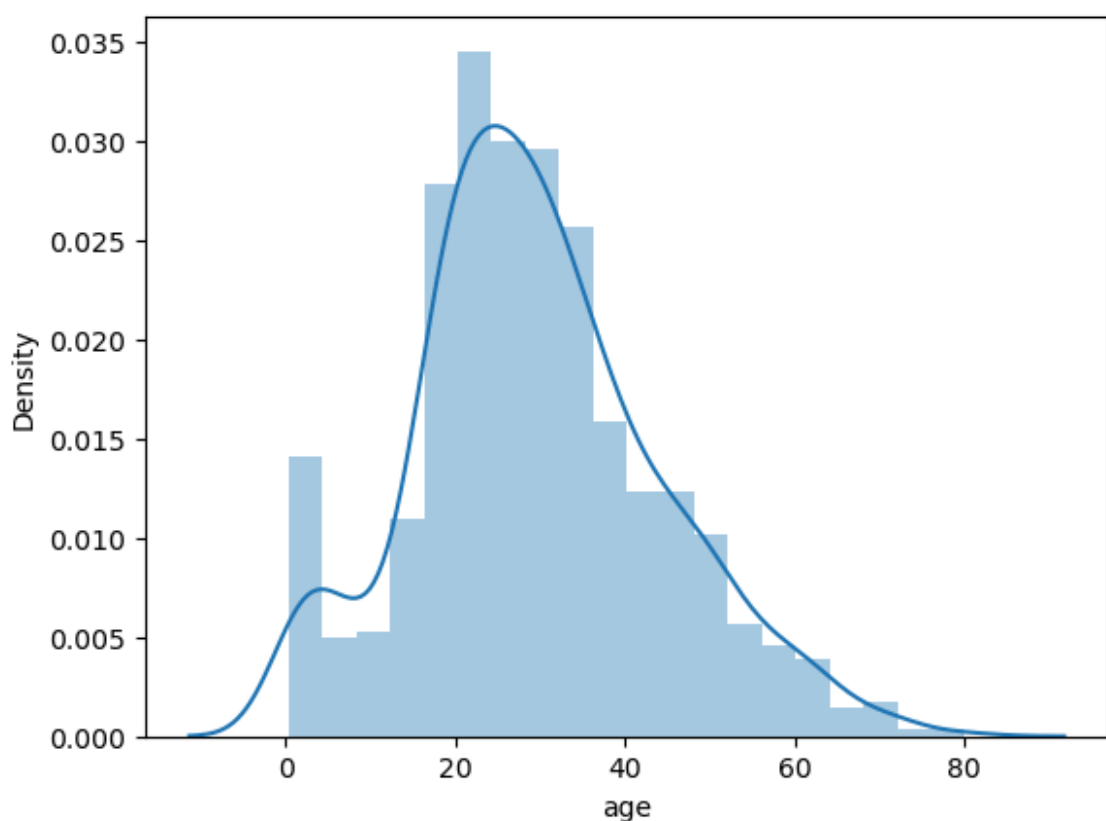C:\Users\alisu\AppData\Local\Temp\ipykernel_10124\3497382289.py:1: UserWarning:

`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 (https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

    sns.distplot(titanic.age)

Out[29]: <Axes: xlabel='age', ylabel='Density'>

In [30]: `sns.distplot(titanic.fare)`

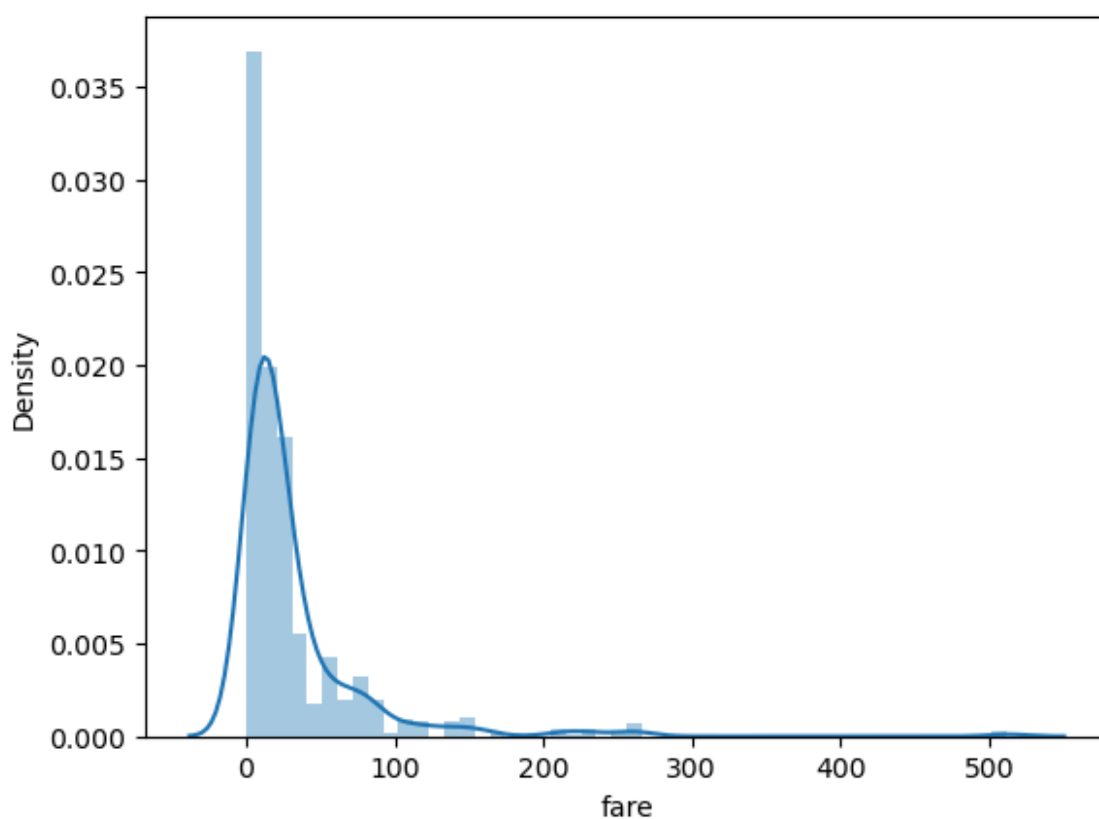C:\Users\alisu\AppData\Local\Temp\ipykernel_10124\3635634896.py:1: UserWarning:

`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).

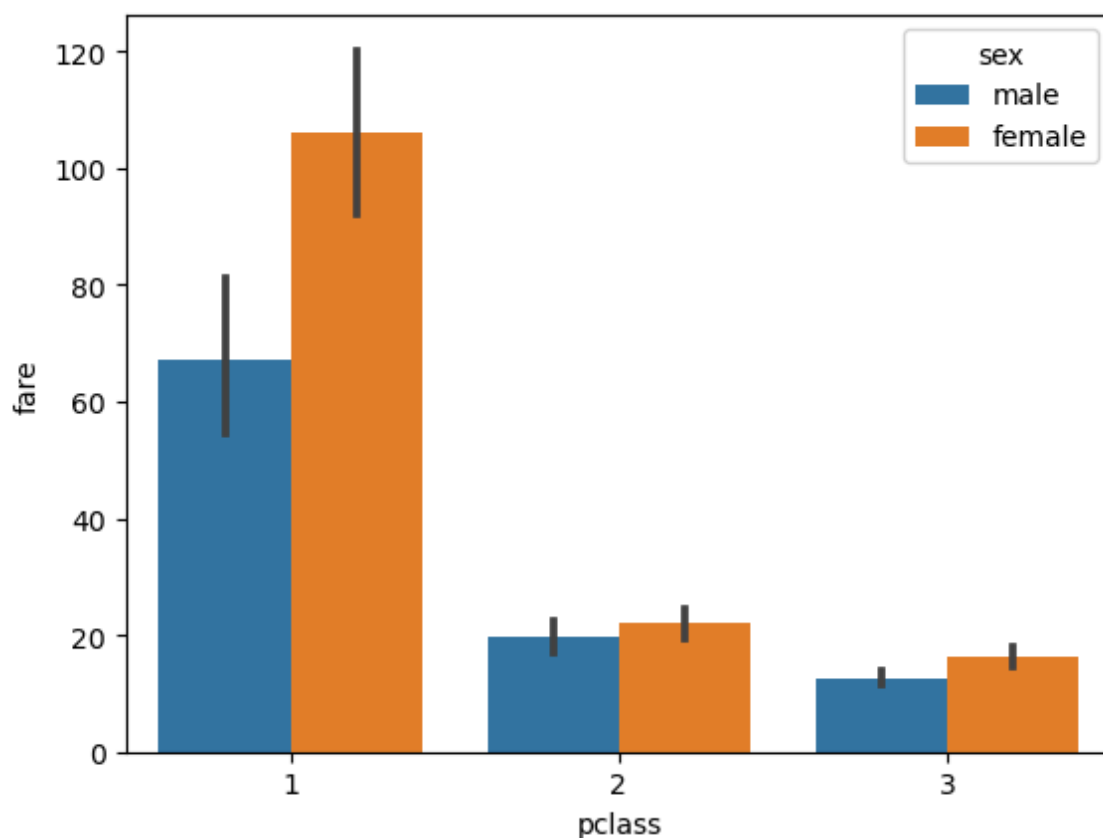For a guide to updating your code to use the new functions, please see https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot(titanic.fare)

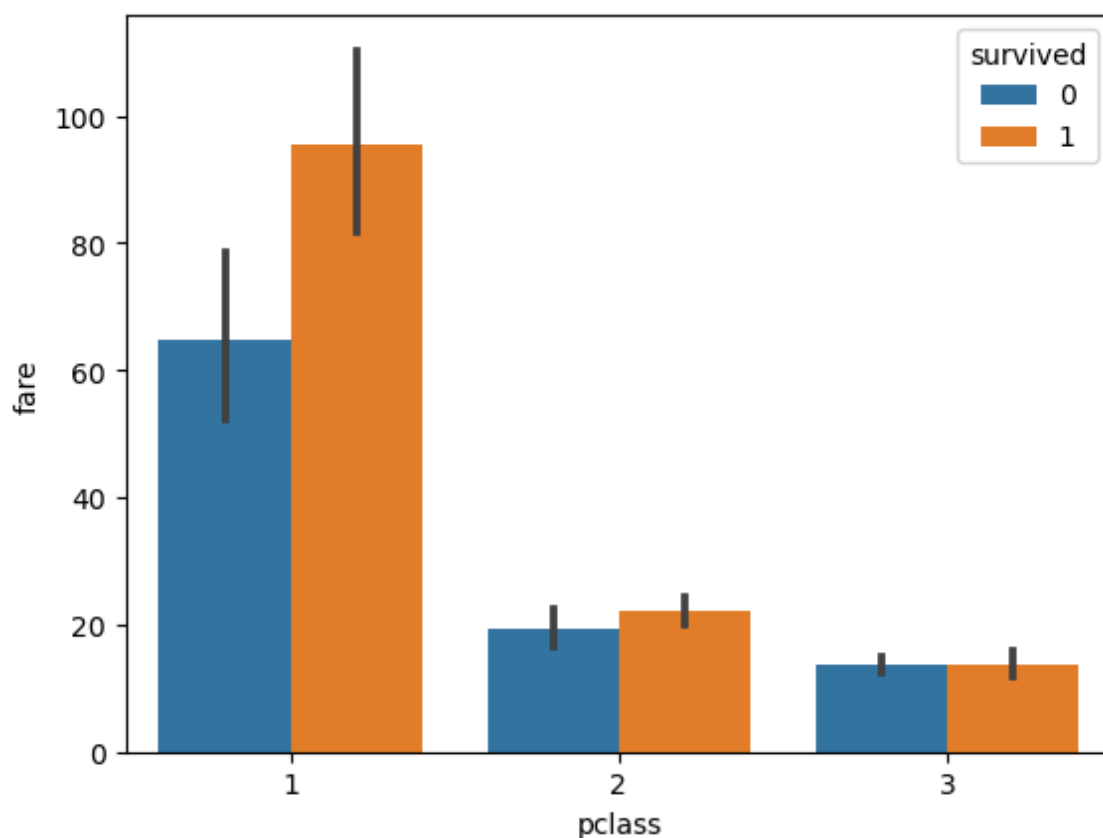Out[30]: <Axes: xlabel='fare', ylabel='Density'>

In [31]: `sns.barplot(x = 'pclass', y = 'fare', data = titanic ,hue = 'sex' )`

Out[31]: `<Axes: xlabel='pclass', ylabel='fare'>`

In [32]: `sns.barplot(x = 'pclass', y = 'fare', data = titanic ,hue = 'survived' )`

Out[32]: `<Axes: xlabel='pclass', ylabel='fare'>`

##Conclusion:
1. THere are more passengers in first that is lower class
2. Age normally distributed
3. Fare is almost left skewed towrards towards
4. The number of old male passenger is more than female passeng
ers
5. There are more young females on ship
6. The females have paid medium to higher ranges of fair.

In [ ]: