pr8

May 4, 2024

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

```python
[17]: data=pd.read_csv("C:\\Users\\nayan\\Downloads\\Titanic-Dataset.csv")
      data.head()
```

```
[17]:    PassengerId  Survived  Pclass  \
     0            1         0       3
     1            2         1       1
     2            3         1       3
     3            4         1       1
     4            5         0       3

                                                       Name     Sex   Age  SibSp  \
     0                            Braund, Mr. Owen Harris    male  22.0      1
     1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
     2                             Heikkinen, Miss. Laina  female  26.0      0
     3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
     4                           Allen, Mr. William Henry    male  35.0      0

        Parch            Ticket     Fare Cabin Embarked
     0      0         A/5 21171   7.2500   NaN        S
     1      0          PC 17599  71.2833   C85        C
     2      0  STON/O2. 3101282   7.9250   NaN        S
     3      0            113803  53.1000  C123        S
     4      0            373450   8.0500   NaN        S
```

```python
[10]: sns.distplot(data['Fare'])
```

C:\Users\nayan\AppData\Local\Temp\ipykernel_18580\2921470011.py:1: UserWarning:
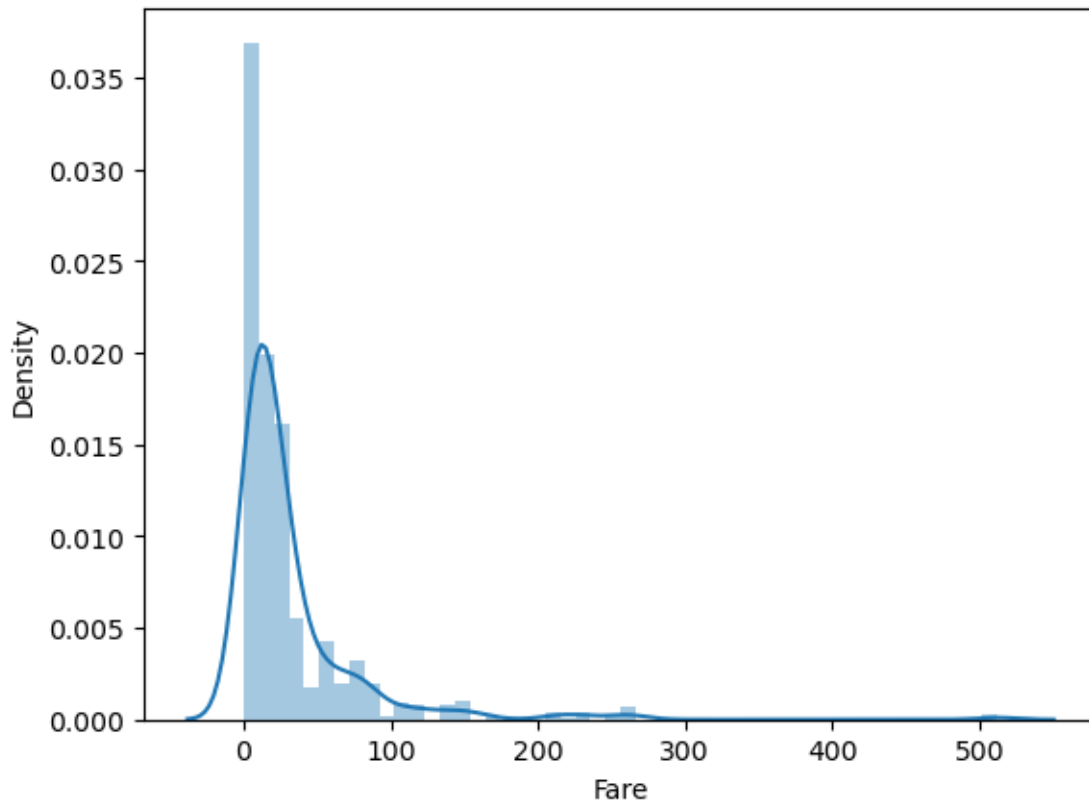
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

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

```
sns.distplot(data['Fare'])
```

[10]: <Axes: xlabel='Fare', ylabel='Density'>

[18]:
```python
sns.distplot(data['Fare'],kde=False)
```

C:\Users\nayan\AppData\Local\Temp\ipykernel_18580\1394462394.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).

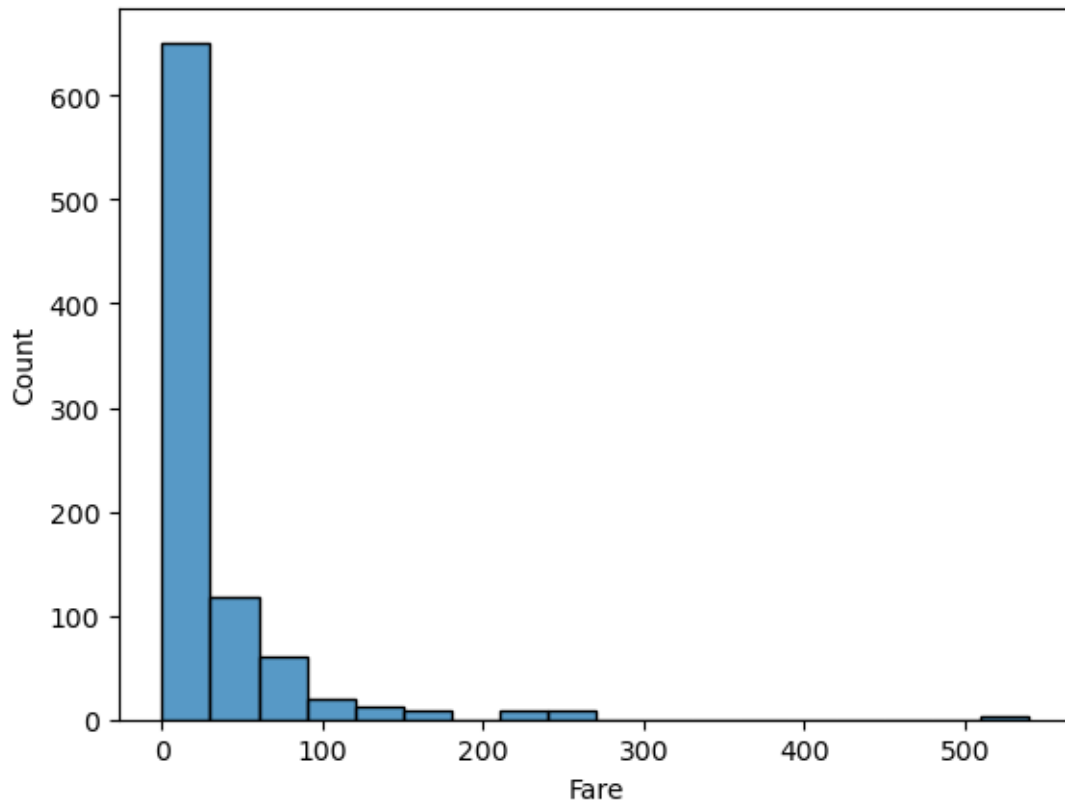For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

```
sns.distplot(data['Fare'],kde=False)
```

[18]: <Axes: xlabel='Fare'>
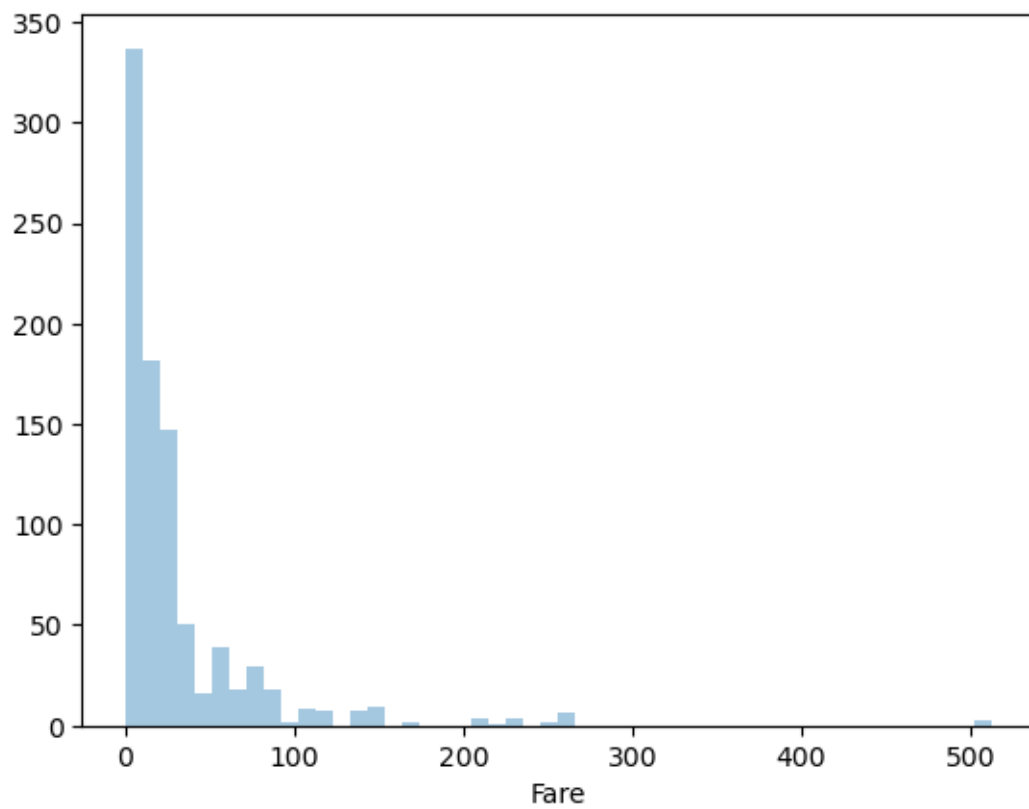
```
[21]: sns.histplot(data=data, x="Fare",binwidth=30)
```

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

[22]: *#Joint Plot It is the combination of the distplot of two variables.It is an* ↪*example of bivariate analysis.We additionally obtain a scatter plot between* ↪*the variable to reflectingtheir linear relationship. We can customize the* ↪*scatter plot into a hexagonal plot, where, more the color intensity, the* ↪*more will be the number of Observation*

```
sns.jointplot(x='Age', y='Fare', data=data)
```
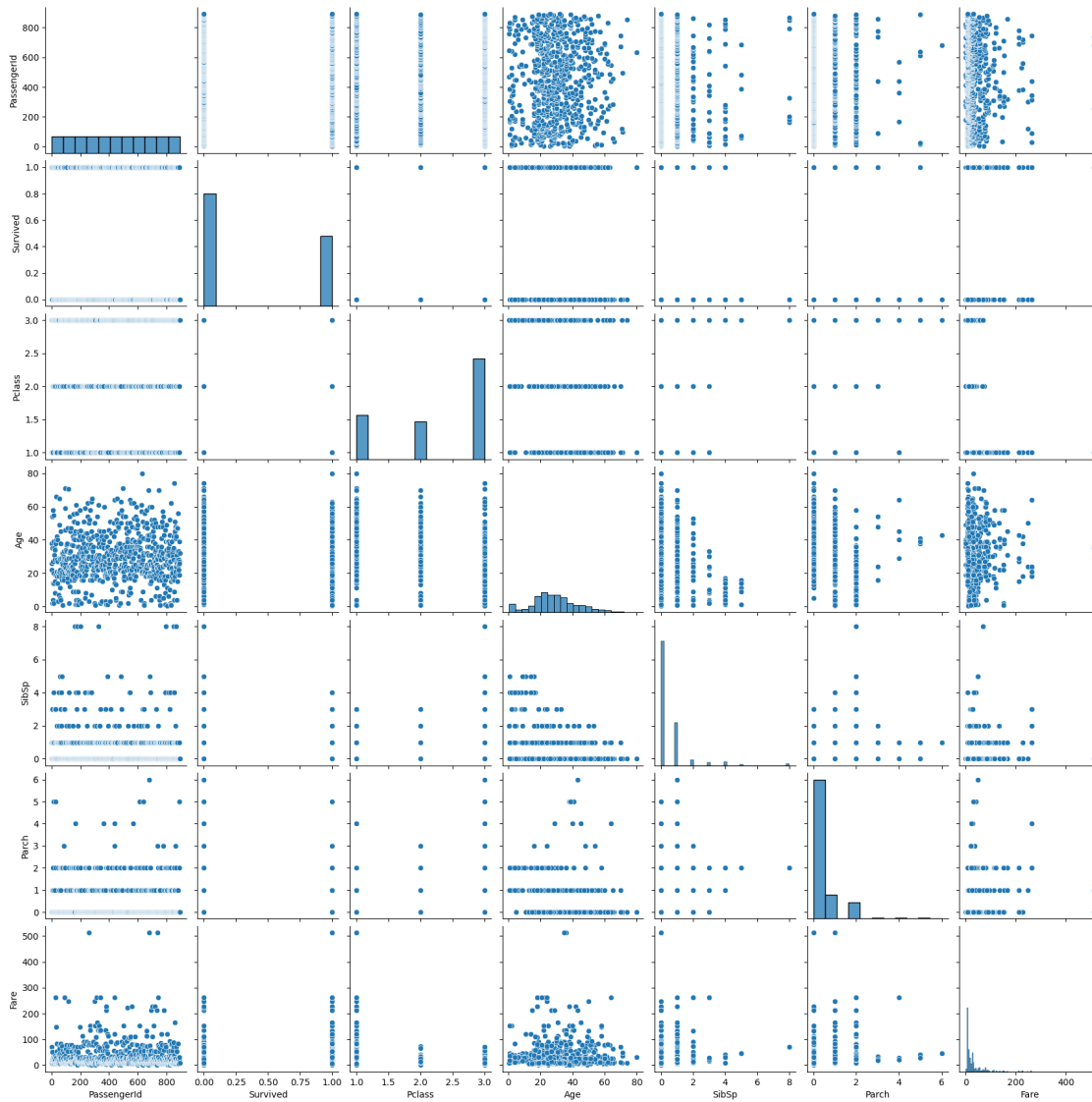
[22]: <seaborn.axisgrid.JointGrid at 0x23a61acc490>
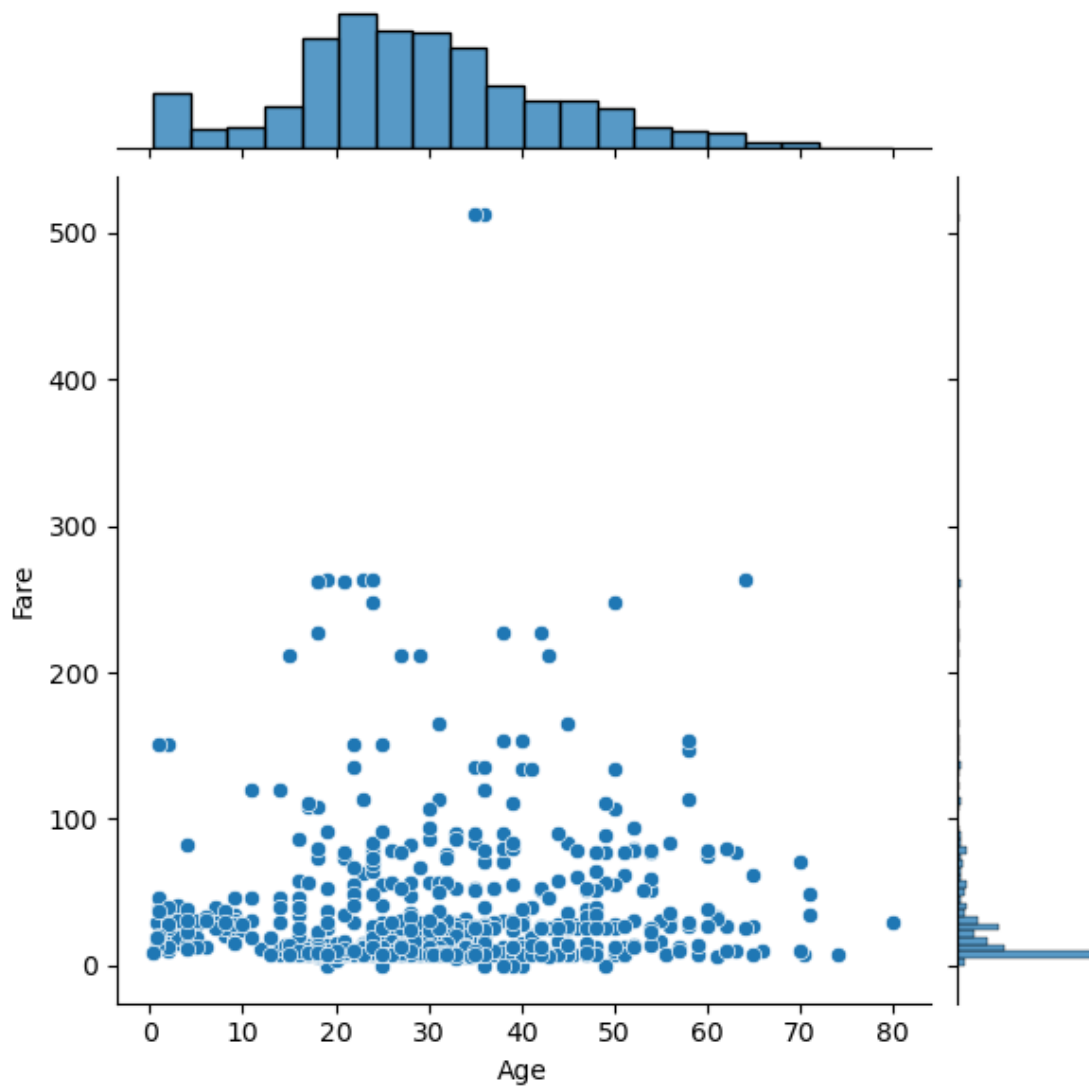
```
[23]: sns.pairplot(data)
```

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

```
[23]: <seaborn.axisgrid.PairGrid at 0x23a60b31cd0>
```

```
[24]: data.describe()
```

```
[24]:        PassengerId    Survived      Pclass         Age       SibSp  \
       count   891.000000  891.000000  891.000000  714.000000  891.000000
       mean    446.000000    0.383838    2.308642   29.699118    0.523008
       std     257.353842    0.486592    0.836071   14.526497    1.102743
       min       1.000000    0.000000    1.000000    0.420000    0.000000
       25%     223.500000    0.000000    2.000000   20.125000    0.000000
       50%     446.000000    0.000000    3.000000   28.000000    0.000000
       75%     668.500000    1.000000    3.000000   38.000000    1.000000
       max     891.000000    1.000000    3.000000   80.000000    8.000000

                   Parch        Fare
```

```
count   891.000000   891.000000
mean      0.381594    32.204208
std       0.806057    49.693429
min       0.000000     0.000000
25%       0.000000     7.910400
50%       0.000000    14.454200
75%       0.000000    31.000000
max       6.000000   512.329200
```

[26]: `data.isnull().sum()`

[26]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

[27]: `sns.distplot(data['Age'])`

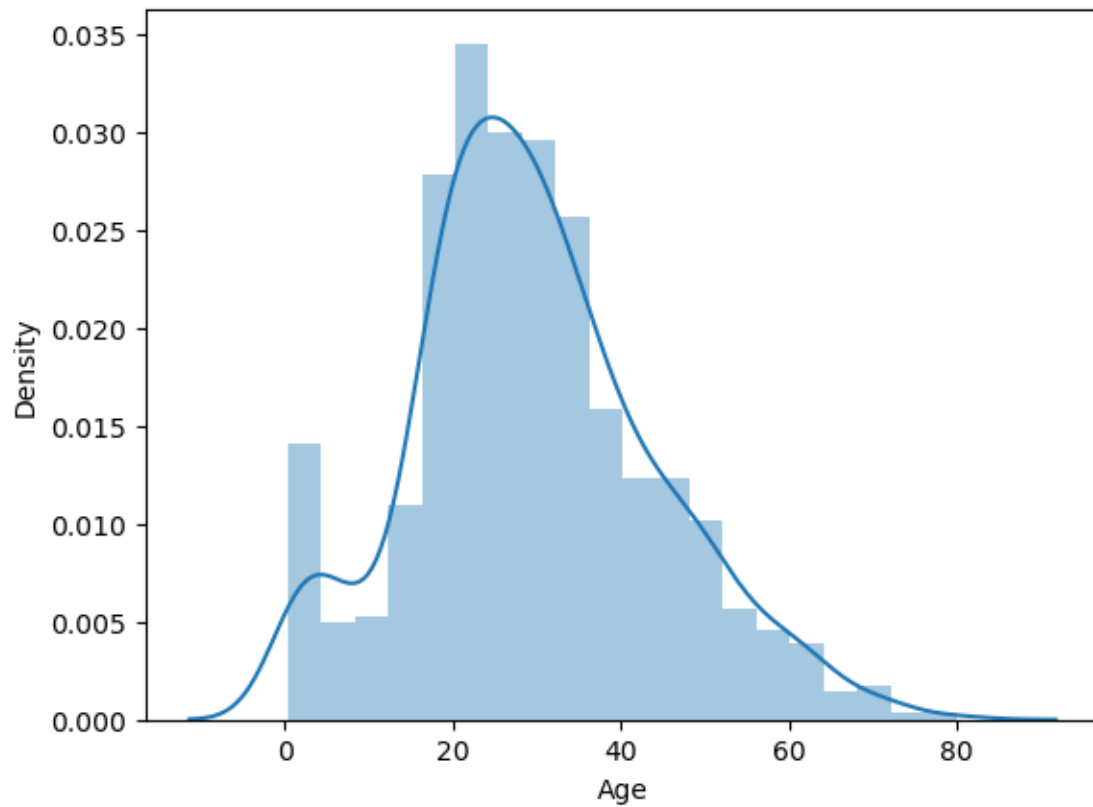C:\Users\nayan\AppData\Local\Temp\ipykernel_18580\2317092479.py:1: UserWarning:

`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

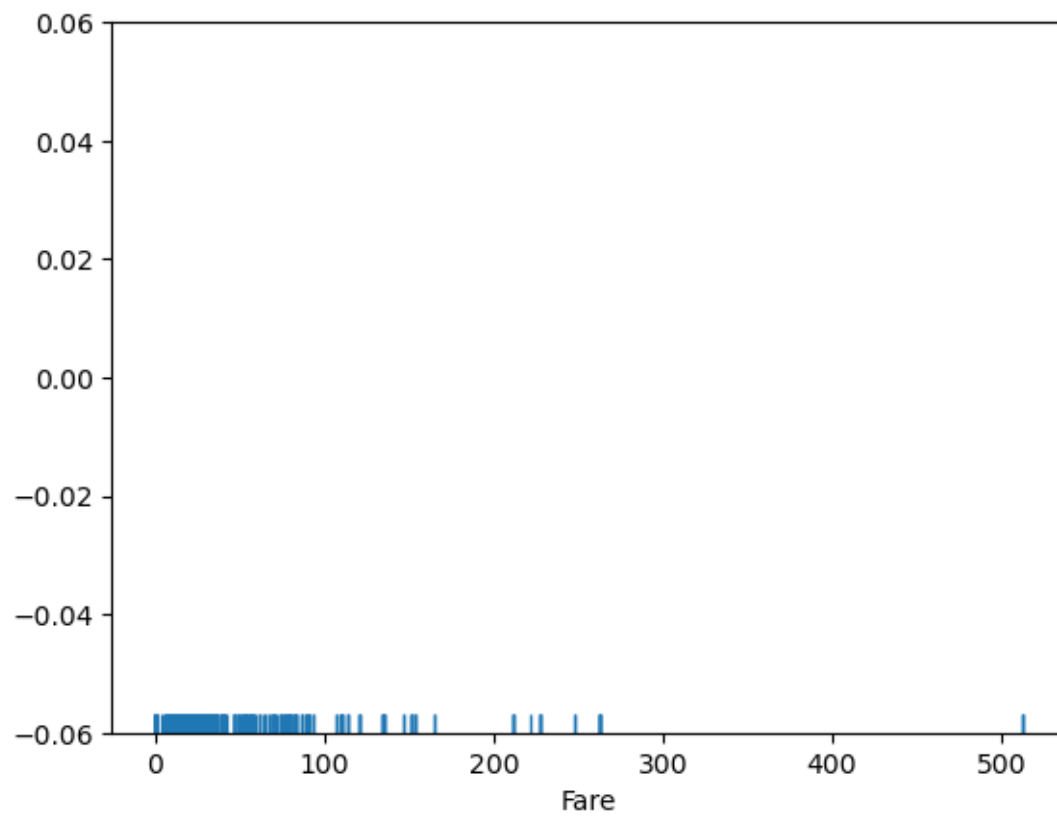 sns.distplot(data['Age'])

[27]: <Axes: xlabel='Age', ylabel='Density'>

[28]: *#Rug Plot It draws a dash mark instead of a uniform distribution as in distplot.*
    ↪ It is an example of a univariate analysis.

```
sns.rugplot(data['Fare'])
```
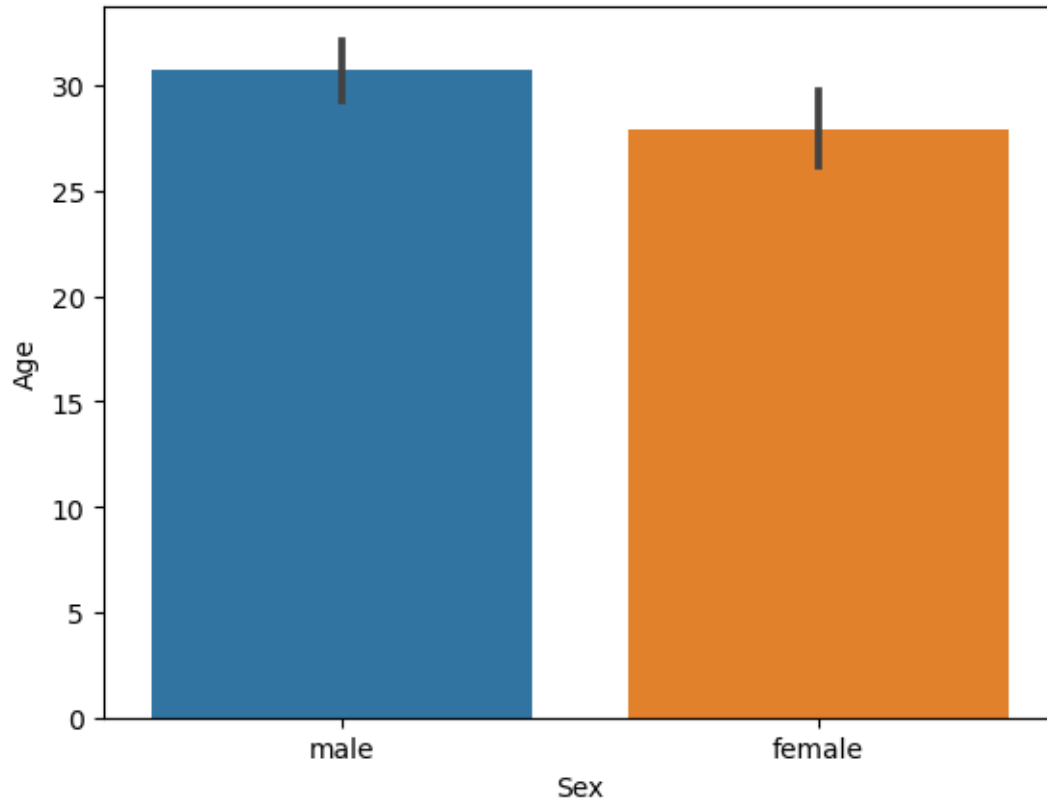
[28]: <Axes: xlabel='Fare'>
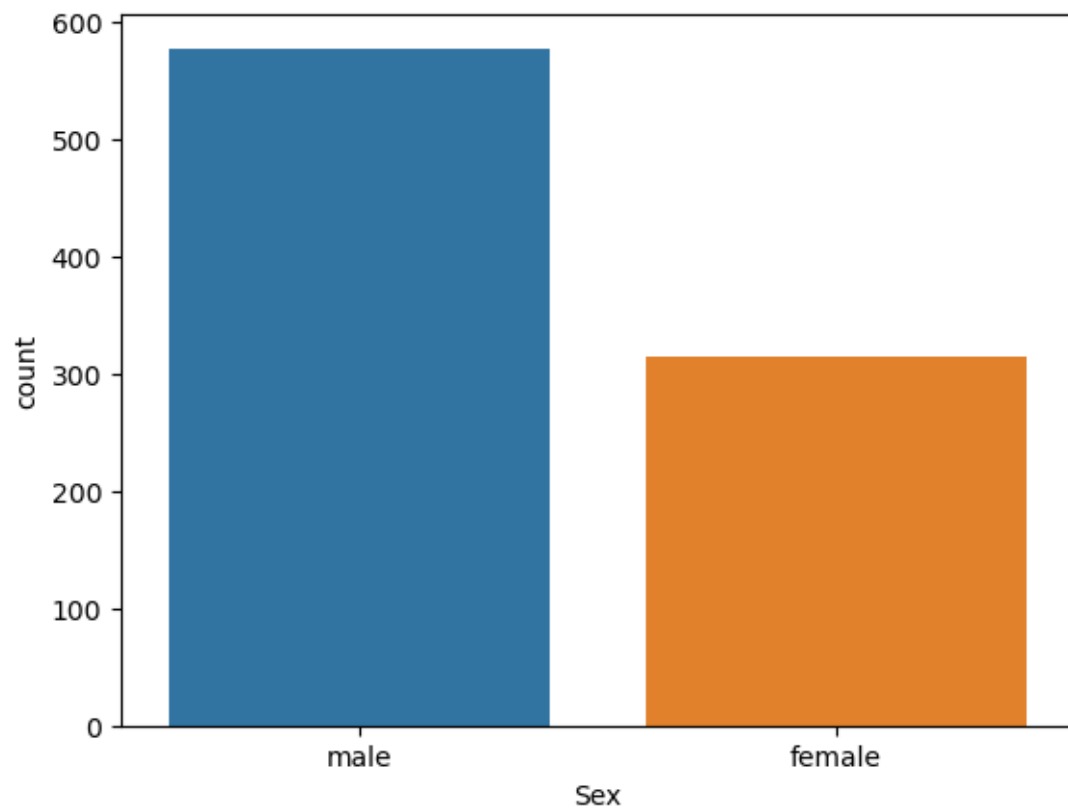
```
[29]: sns.barplot(x='Sex',y='Age',data=data)
```

```
[29]: <Axes: xlabel='Sex', ylabel='Age'>
```

[30]: *##Count Plot It counts the number of occurrences of categorical variables.It is↵*
↪an example of a univariate analysis.The count plot is similar to the bar↵
↪plot, however itdisplays the count of the categories in a specific column.↵
↪For instance, if we want to count the number of males and women passenger we↵
↪can do so using count plot asfollows:

```
sns.countplot(x='Sex',data=data)
```

[30]: `<Axes: xlabel='Sex', ylabel='count'>`

[ ]: