```
    Assignment 8Data Visualization I

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

    Assignment 9Data Visualization II

    1. Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box plot for distribution of
age with respect to each gender along with the information about whether

    they survived or not. (Column names : 'sex' and 'age') Write observations on the inference from the above
statistics.
```

In [5]:

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
dataset=pd.read csv("titanic.csv")
```

In [6]:

```
dataset
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [7]:

```
dataset.head()
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

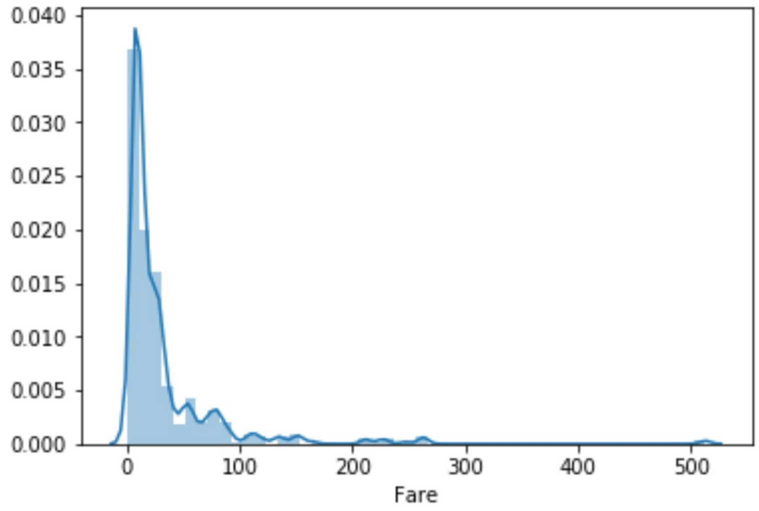
Distribution Plots

```
a. Distplot
b. jointplot
```

#TO FIND THE DISTRIBUTION OF FARE COLUMN USING HISTOGRAM WE USE Distplot

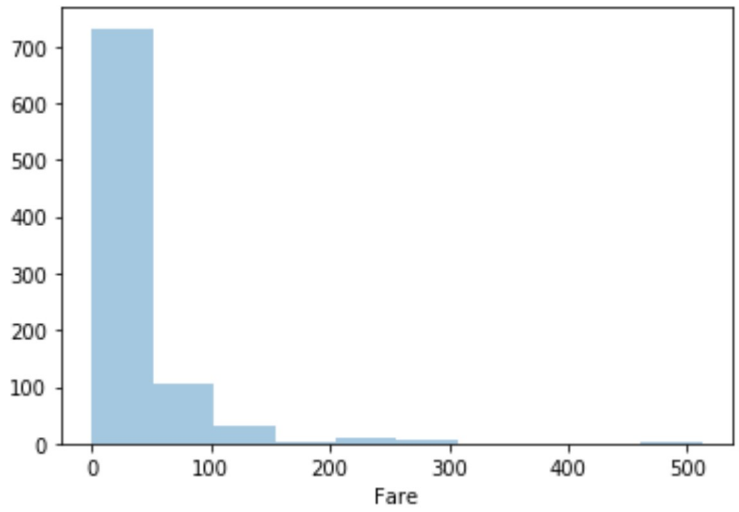
In [8]:
```
sns.distplot(dataset['Fare'])
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x21434582648>
```



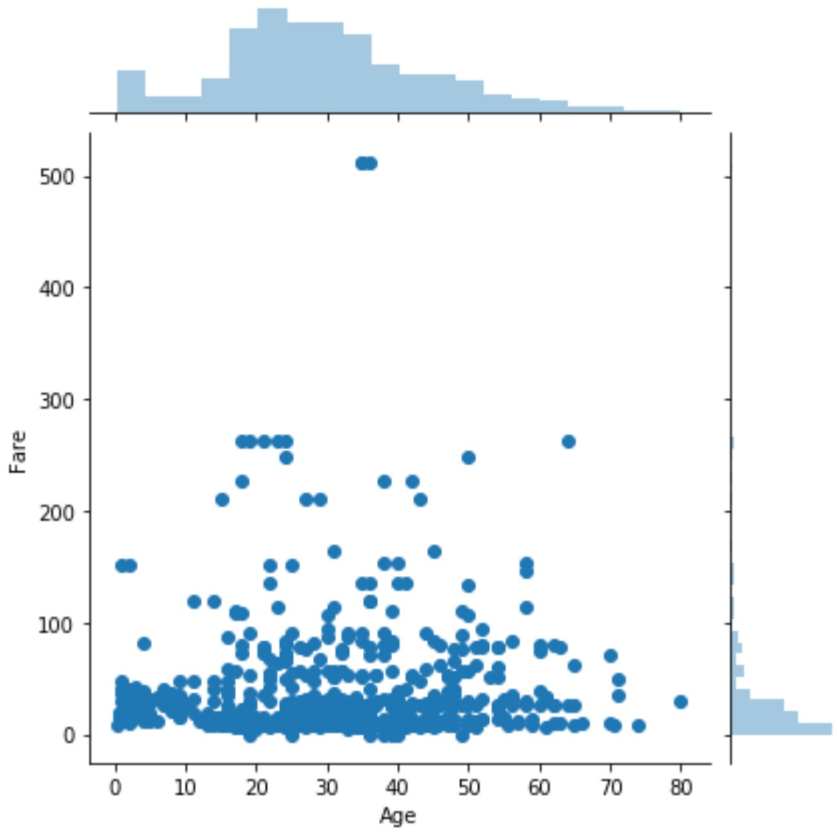
In [9]:
```
## increase the bin size

sns.distplot(dataset['Fare'],kde=False,bins=10)
```

Out[9]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x21434d79208>
```



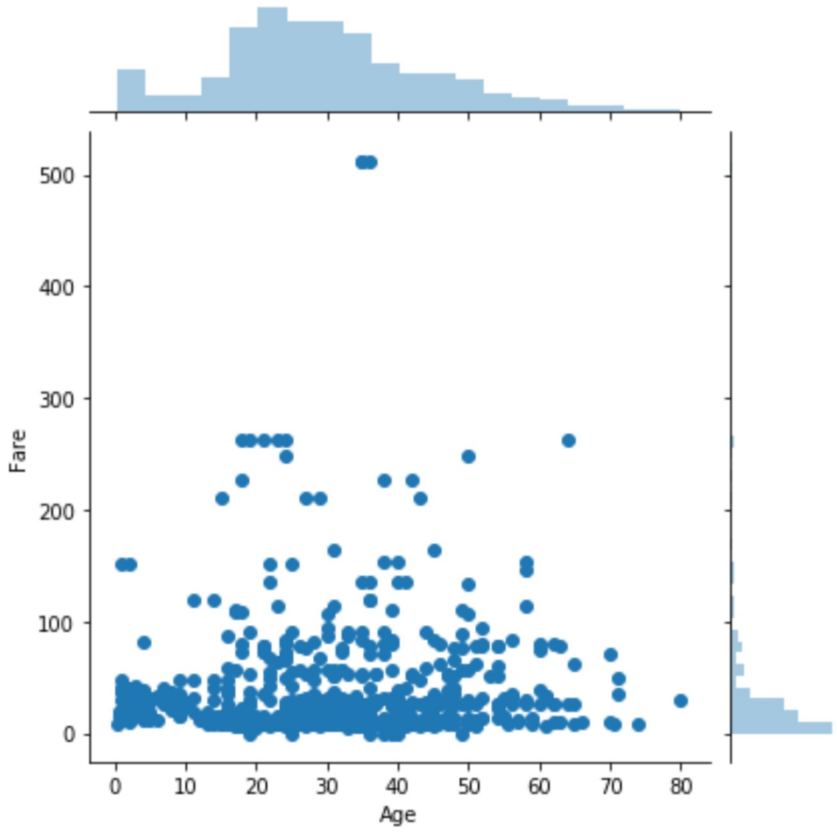
In [10]:
```
## joinplot works for 2 variable byvarient analysis obtain a scatter plot between the variable
sns.jointplot(x='Age',y='Fare',data=dataset)
```

Out[10]:
```
<seaborn.axisgrid.JointGrid at 0x21434e13d08>
```



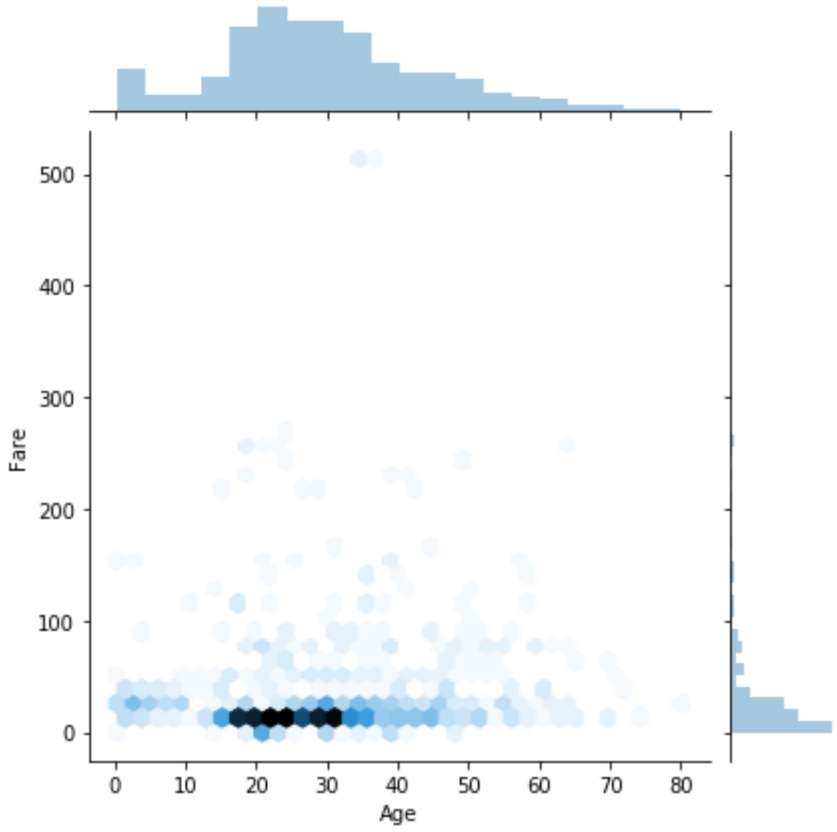
```
In [15]: sns.jointplot(x=dataset['Age'],y=dataset['Fare'],kind='scatter')
```

```
Out[15]: <seaborn.axisgrid.JointGrid at 0x2143525d708>
```



```
In [18]: sns.jointplot(x=dataset['Age'],y=dataset['Fare'],kind='hex')
```

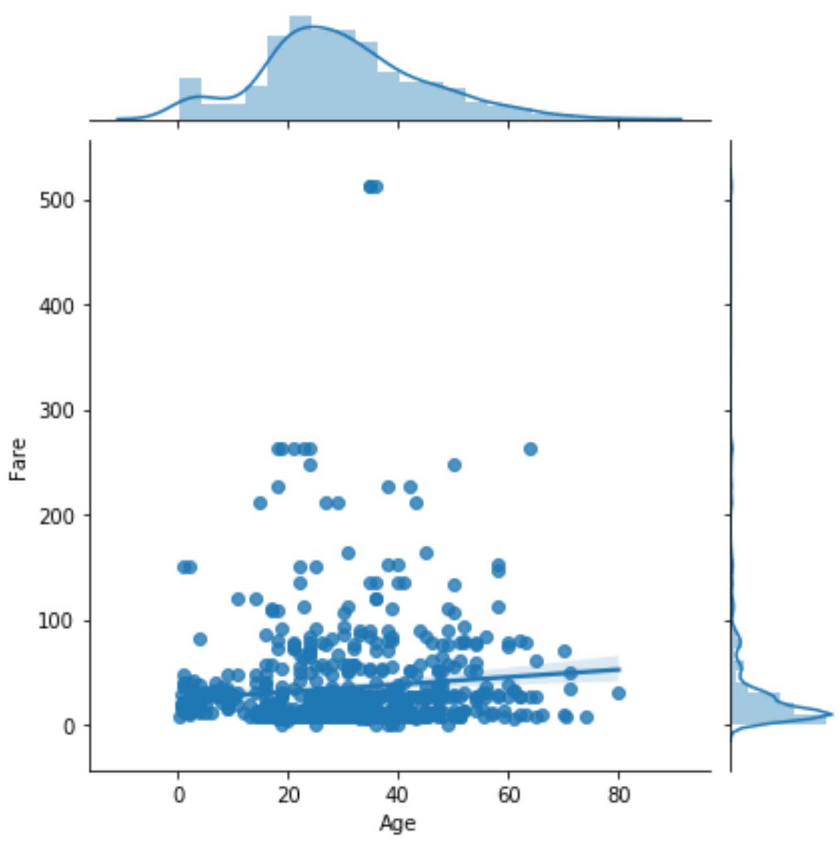
```
Out[18]: <seaborn.axisgrid.JointGrid at 0x214351fb508>
```



We can see that there no appropriate linear relation between age and fare. kind = ‘hex’ provides the hexagonal plot and kind = ‘reg’ provides a regression line on the graph.

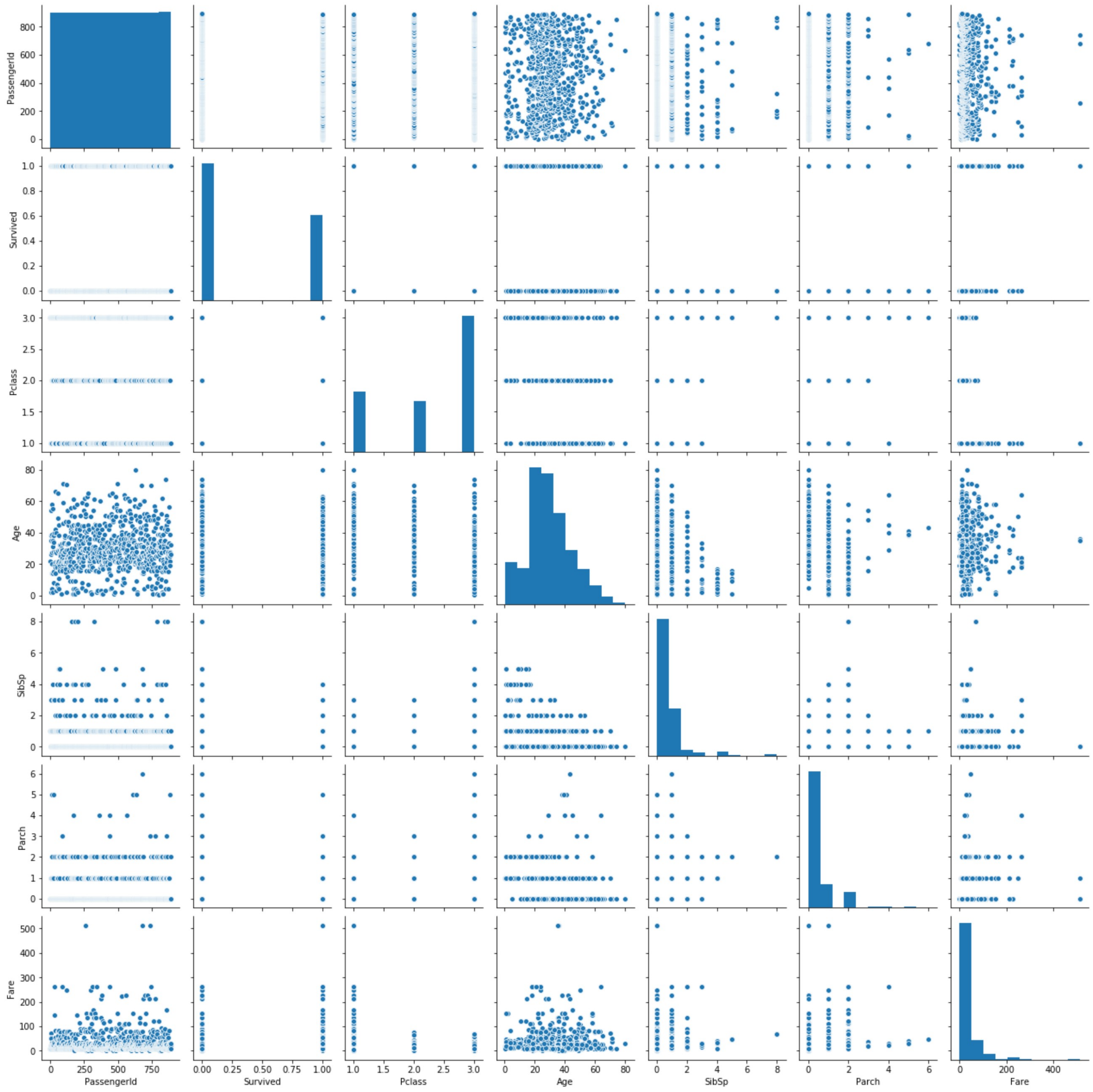
In [20]:
```
sns.jointplot(x=dataset['Age'],y=dataset['Fare'],kind='reg')  # no linear relation
```

Out[20]:
```
<seaborn.axisgrid.JointGrid at 0x2143567e708>
```



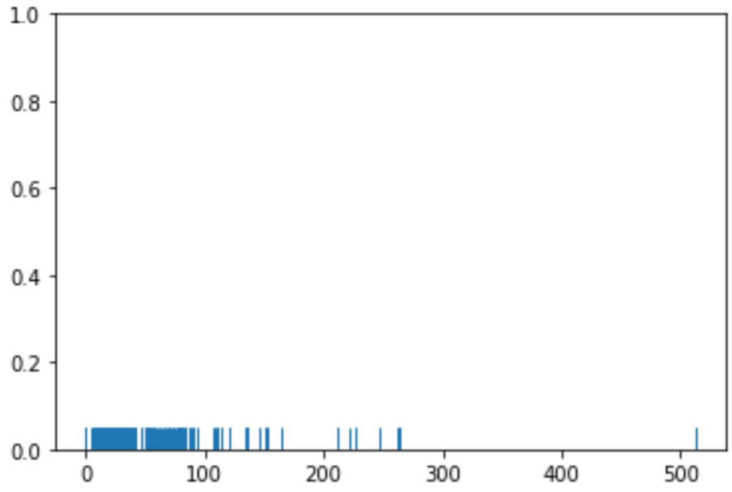
In [21]:
```
sns.pairplot(dataset)
```

Out[21]:
```
<seaborn.axisgrid.PairGrid at 0x214351baec8>
```



In [27]:
```
# it similar to distplot
sns.rugplot(dataset['Fare'])
```

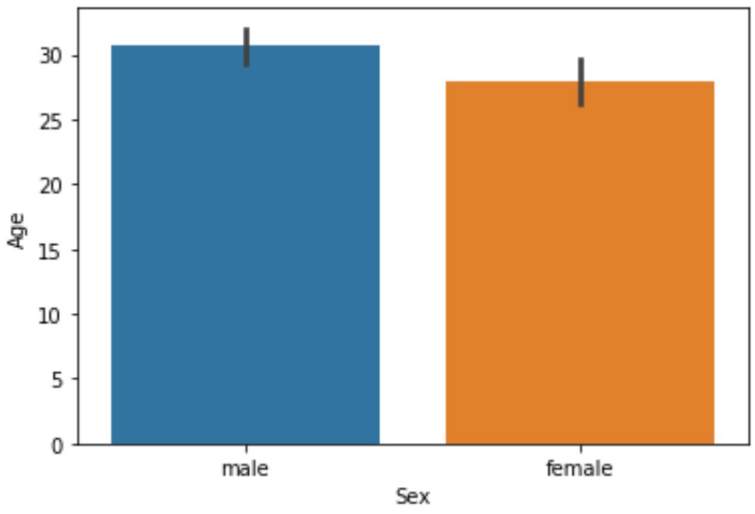
Out[27]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x214387f1c08>
```



```
## categorical plots for categorical variables
Categorical plots, as the name suggests are normally used to plot categorical data. The categorical plots
plot the values in the categorical column against another categorical column or a numeric column
```

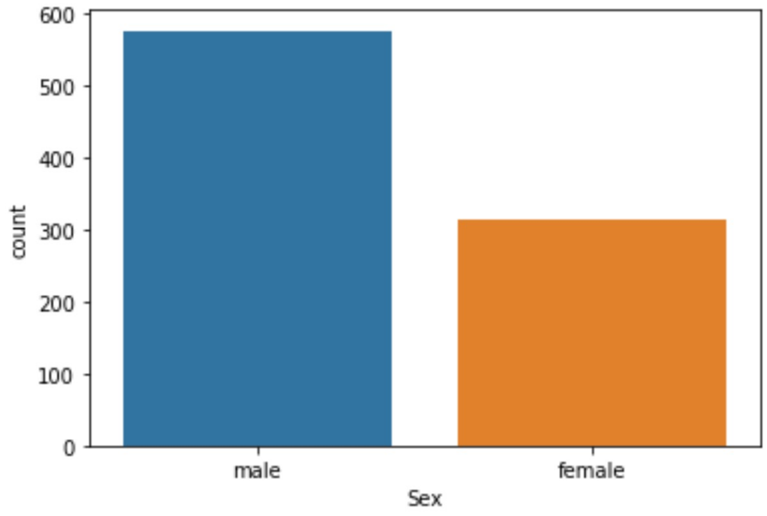
In [29]:
```
# a.Bar plot
sns.barplot(x=dataset['Sex'],y=dataset['Age'])
```

Out[29]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x214388ad648>
```



In [32]:
```
# b.Count plot for a single varible
sns.countplot(x='Sex',data=dataset)
```

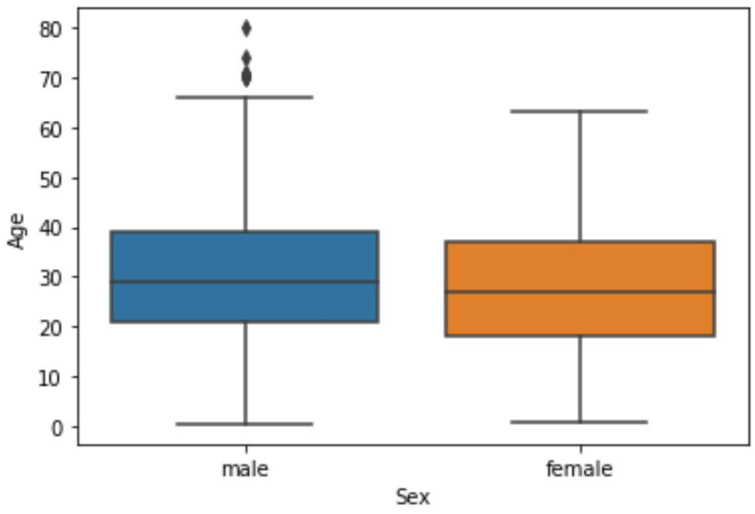
Out[32]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2143896a348>
```



```
##c. Box Plot It is a 5 point summary plot. It gives the information about the maximum, minimum, mean, first
quartile, and third quartile of a continuous variable. Also, it equips us with knowledge of outliers. We can
plot this for a single continuous variable or can analyze different categorical variables based on a
continuous variable.
```

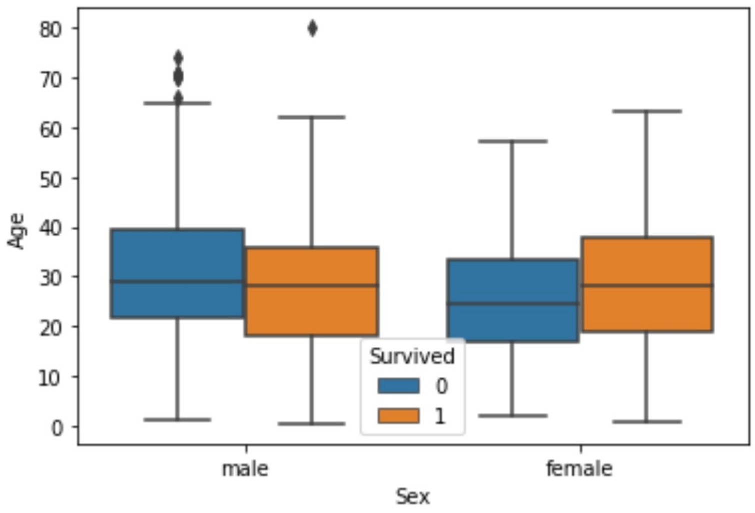
```
In [33]: sns.boxplot(x='Sex',y='Age',data=dataset)
```

```
Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0x21438b6db88>
```



```
In [36]: # using hue
         sns.boxplot(x='Sex',y='Age',data= dataset,hue='Survived')
```

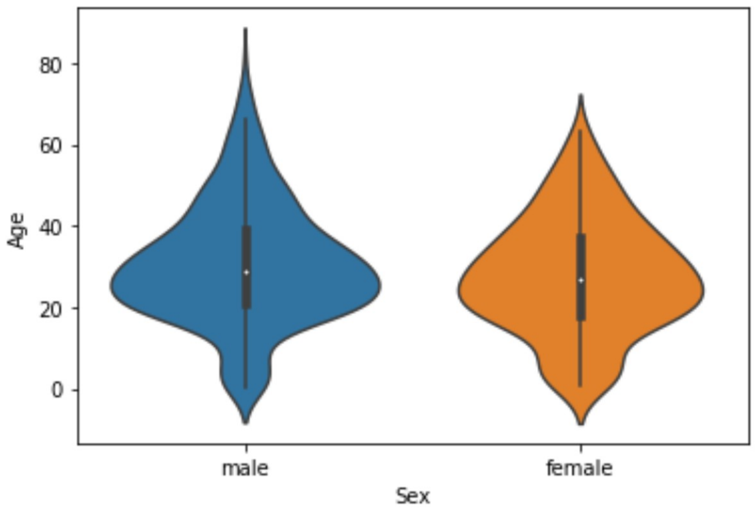
```
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x21438cd5bc8>
```



d.VIOLIN PLOT The violin plot is similar to the box plot, however, the violin plot allows us to display all the components that actually correspond to the data point. The violinplot() function is used to plot the violin plot. Like the box plot, the first parameter is the categorical column, the second parameter is the numeric column while the third parameter is the dataset.

```
In [38]: sns.violinplot(x='Sex',y='Age',data=dataset)
```

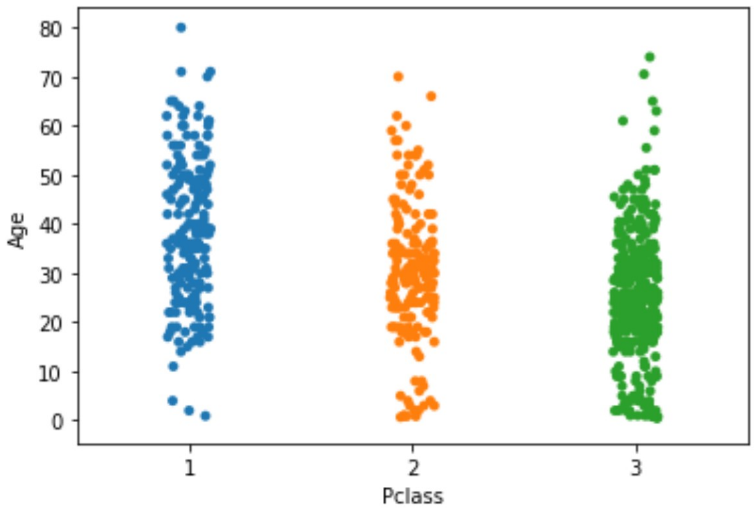
```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x21438cdac88>
```



```
In [39]: # advance plots
         #1. strip plot
```

```
In [40]: sns.stripplot(y = dataset['Age'], x = dataset['Pclass'])
```

```
Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x21438f9b248>
```

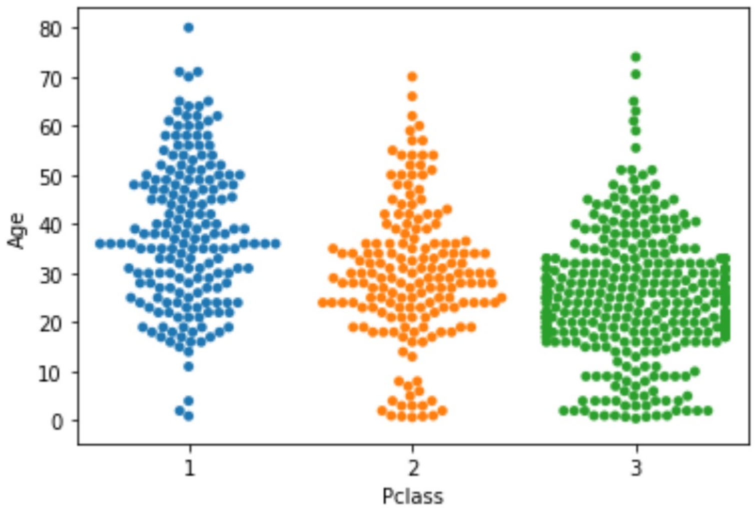


```
# swarn plot:

    b. Swarm Plot
    It is the combination of a strip plot and a violin plot.

    Along with the number of data points, it also provides their respective distribution.
```

```
In [42]: sns.swarmplot(y = dataset['Age'], x = dataset['Pclass'])
```

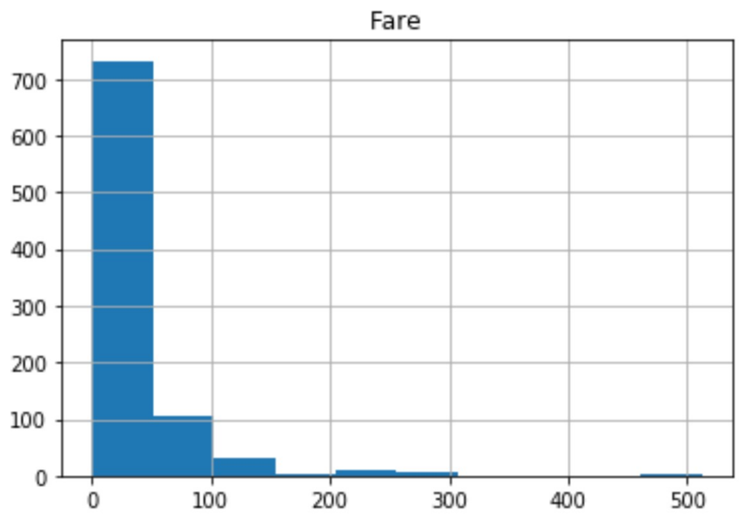
```
C:\Users\ADMIN\anaconda3\lib\site-packages\seaborn\categorical.py:1326: RuntimeWarning: invalid value encoun
tered in less
  off_low = points < low_gutter
C:\Users\ADMIN\anaconda3\lib\site-packages\seaborn\categorical.py:1330: RuntimeWarning: invalid value encoun
tered in greater
  off_high = points > high_gutter
```

```
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x214390bd648>
```



```
In [43]: dataset.hist('Fare')
```

```
Out[43]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000214389933C8>]],
               dtype=object)
```



```
In [ ]:
```