In [ ]:
```
NAME : ARYAN SIRDESAI
ROLL NO. : TACO20175
Lab Assignment 8 : Data Visualization I

Problem Statement :
1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains i
passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to se
patterns in the data.
2. Write a code to check how the price of the ticket (column name: 'fare') for eac
plotting a histogram.
```

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

Loading "titanic" dataset

In [2]:
```
df = pd.read_csv('titanic.csv')
df
```

Out[2]:

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

891 rows × 12 columns

In [3]:
```
df.columns
```

Out[3]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [4]:
```
df.describe()
```

Out[4]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
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
```

GroupBy "Survived" Attribute by count.

In [6]:
```
df_c= df.groupby(by='Survived').agg(count=('PassengerId','count'))
```

In [7]:
```
df_c
```

Out[7]:

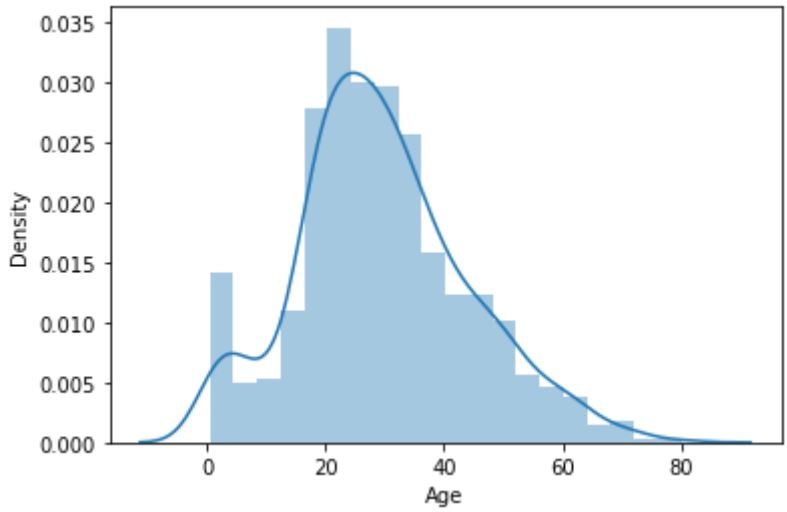
Survived	count
0	549
1	342

Survived value - 0 indicates dead and 1 indicates survived !

In [8]:
```
sns.distplot(df['Age'])
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[8]:
```
<AxesSubplot:xlabel='Age', ylabel='Density'>
```

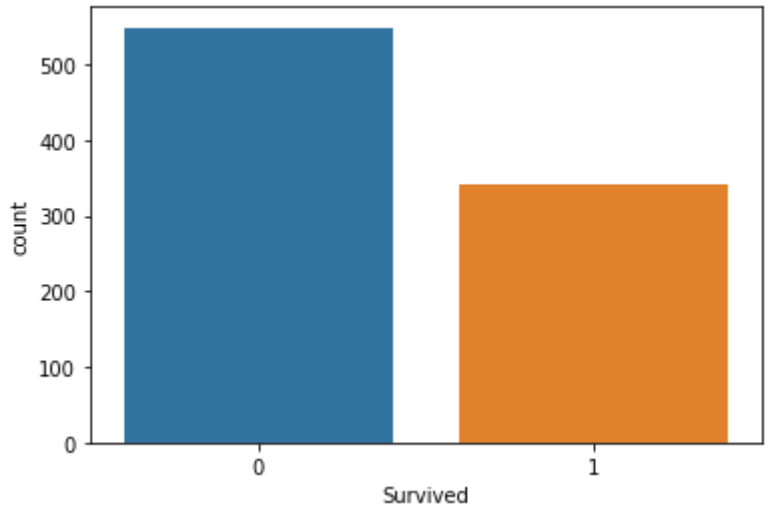


The plot indicates there are more number of passengers of age group between 20 to 40 !

In [9]:
```
sns.countplot(df['Survived'])
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWar
ning: Pass the following variable as a keyword arg: x. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an e
xplicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

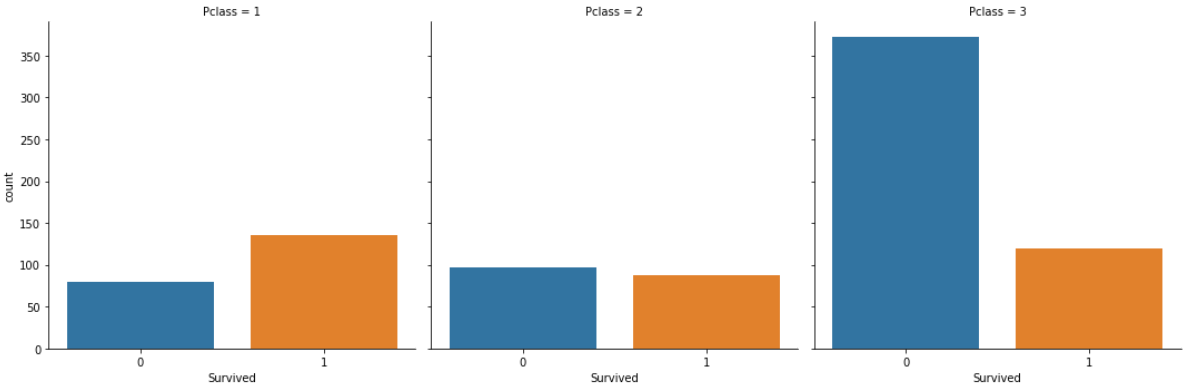
Out[9]:
```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```



Pclass vs Survived

In [10]:
```
sns.catplot(x ="Survived",col='Pclass',
kind ="count", data = df)
```

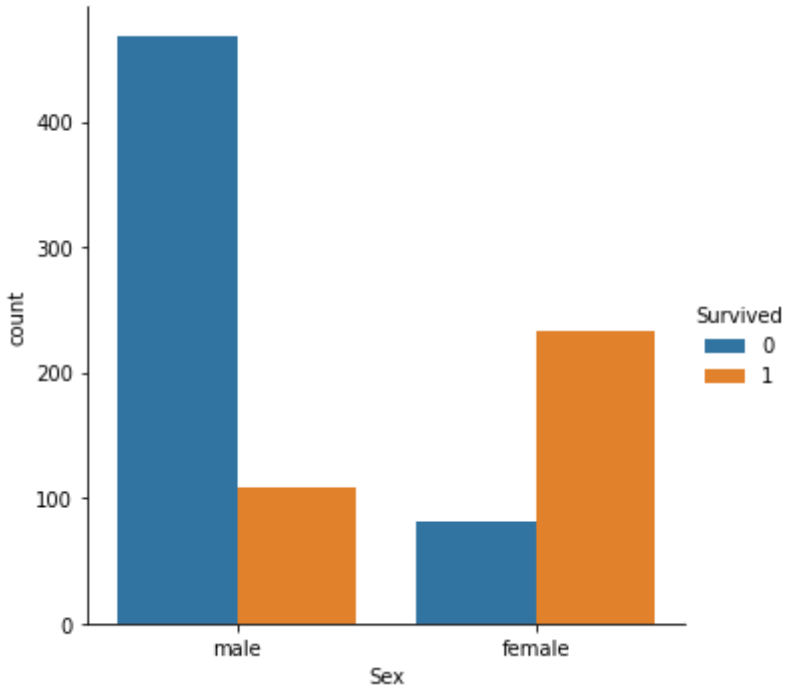
Out[10]:
```
<seaborn.axisgrid.FacetGrid at 0x7f3e8e5e00d0>
```



Gender vs Survived

In [11]:
```
sns.catplot(x ="Sex", hue ="Survived",
kind ="count", data = df)
```

Out[11]:
```
<seaborn.axisgrid.FacetGrid at 0x7f3e8e4acd30>
```

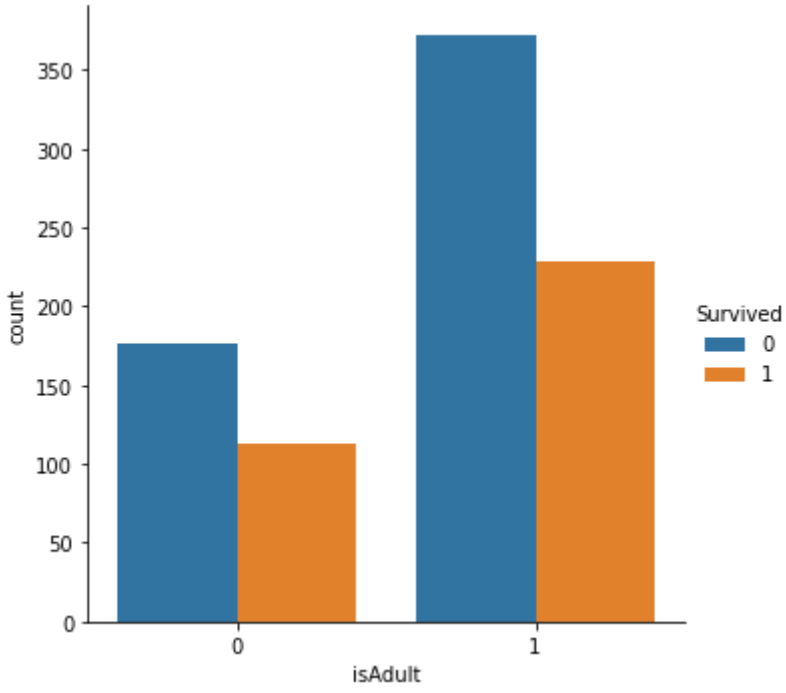


Survived vs AgeGroup (Adult / Non-Adult)

In [12]:
```
df['isAdult'] = np.where(df['Age']>=18,1,0)
```

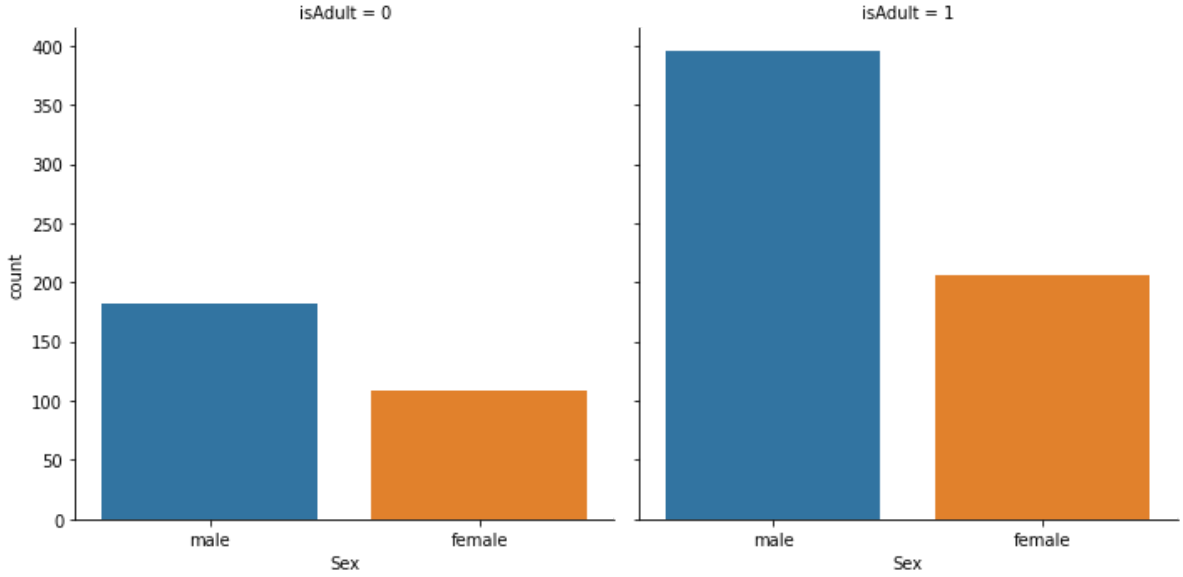
In [13]:
```
sns.catplot(x ="isAdult", hue ="Survived",
kind ="count", data = df)
```

Out[13]:
```
<seaborn.axisgrid.FacetGrid at 0x7f3e8e656700>
```



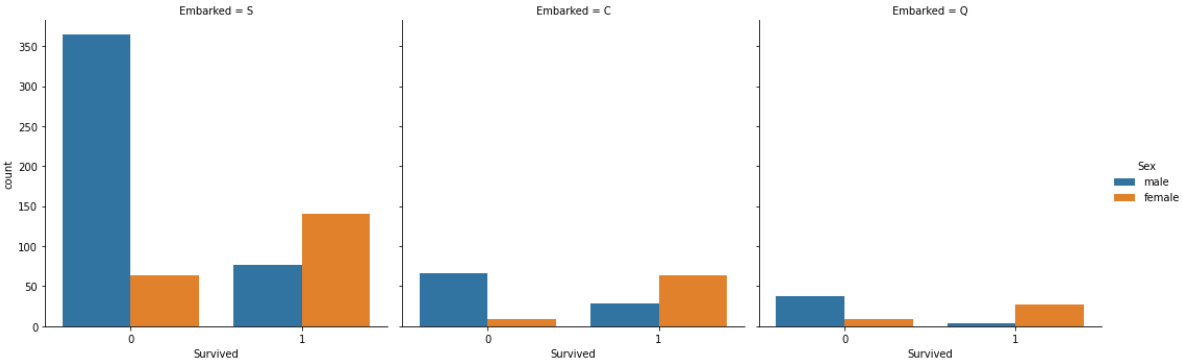
In [14]:
```
sns.catplot(x ="Sex",col='isAdult',
kind ="count", data = df)
```

Out[14]:
```
<seaborn.axisgrid.FacetGrid at 0x7f3e8e0ae040>
```



In [22]:
```
sns.catplot(x ="Survived",col='Embarked', hue='Sex',
kind ="count", data = df)
```

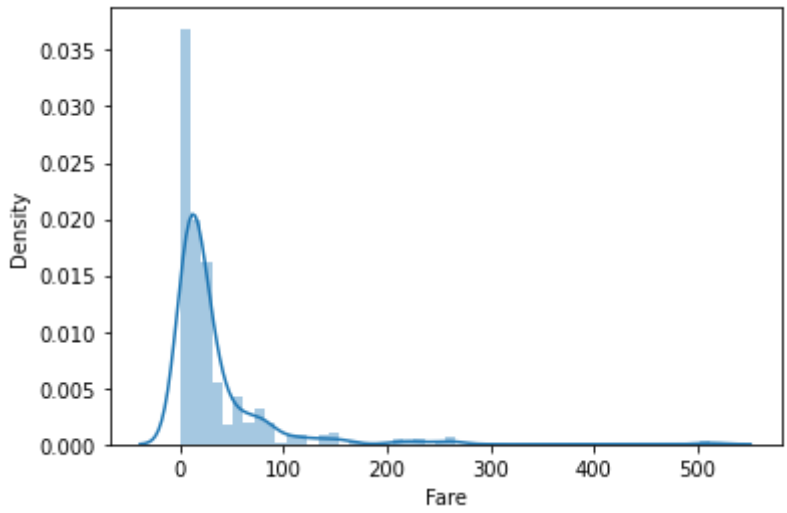
Out[22]:
```
<seaborn.axisgrid.FacetGrid at 0x7f3e8e0ae0a0>
```



In [20]:
```
sns.distplot(df['Fare'])
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[20]:
```
<AxesSubplot:xlabel='Fare', ylabel='Density'>
```



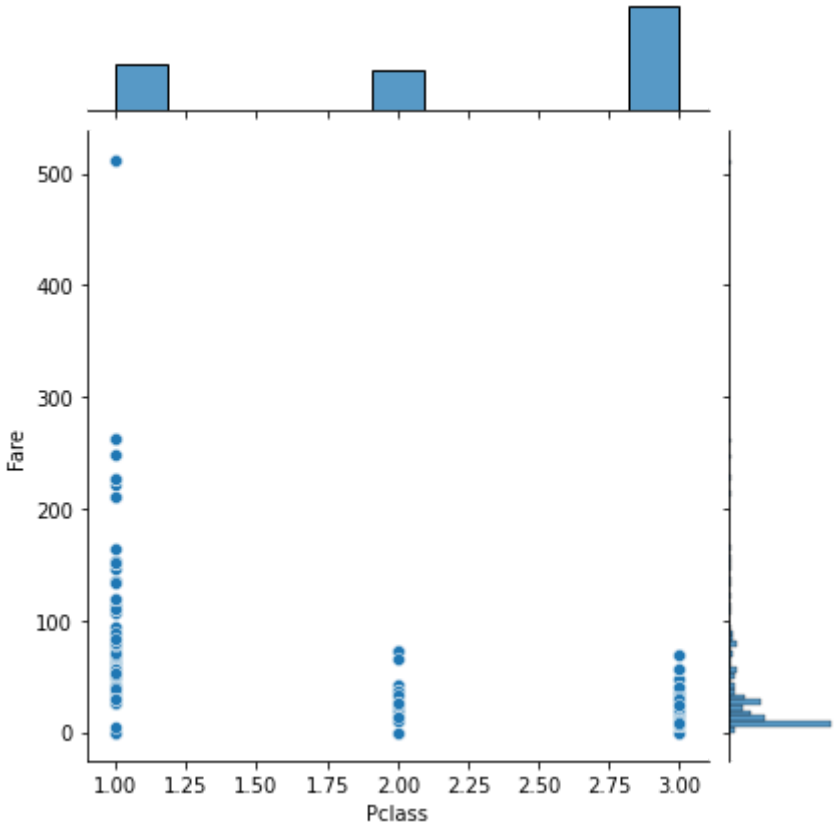
In [16]:
```
df.corr()
```

Out[16]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	isAdu
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658	0.0348
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307	-0.0083
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500	-0.2364
Age	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067	0.6170
SibSp	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651	-0.2463
Parch	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225	-0.1199
Fare	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000	0.0911
isAdult	0.034839	-0.008309	-0.236475	0.617063	-0.246303	-0.119937	0.091114	1.0000

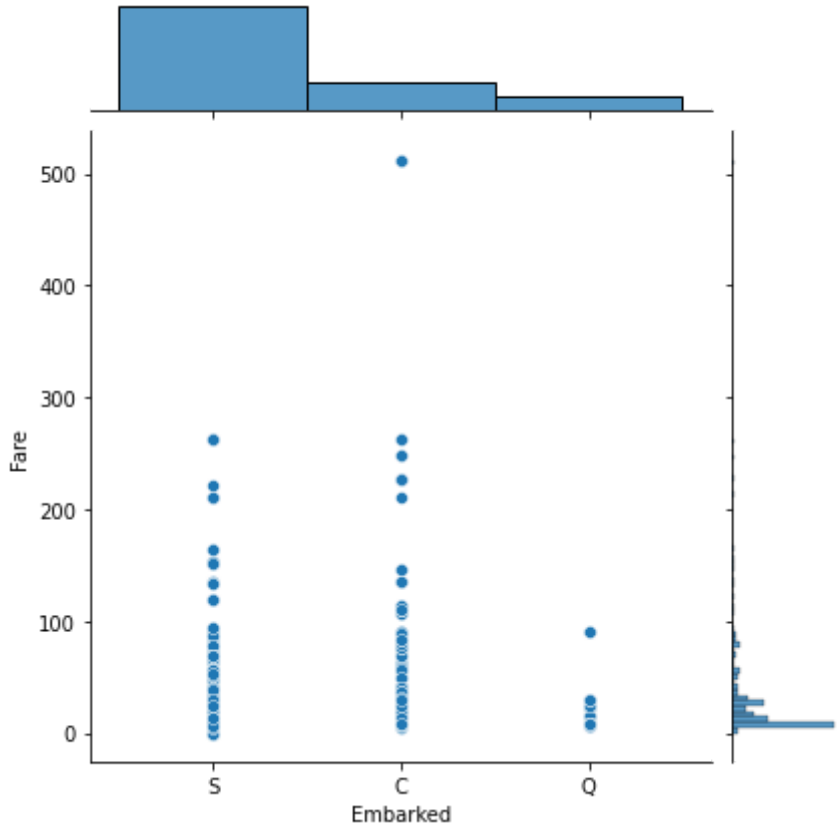
In [17]:
```
sns.jointplot(x = "Pclass", y = "Fare",
              kind = "scatter", data = df)
```

Out[17]:
```
<seaborn.axisgrid.JointGrid at 0x7f3e8dd407c0>
```



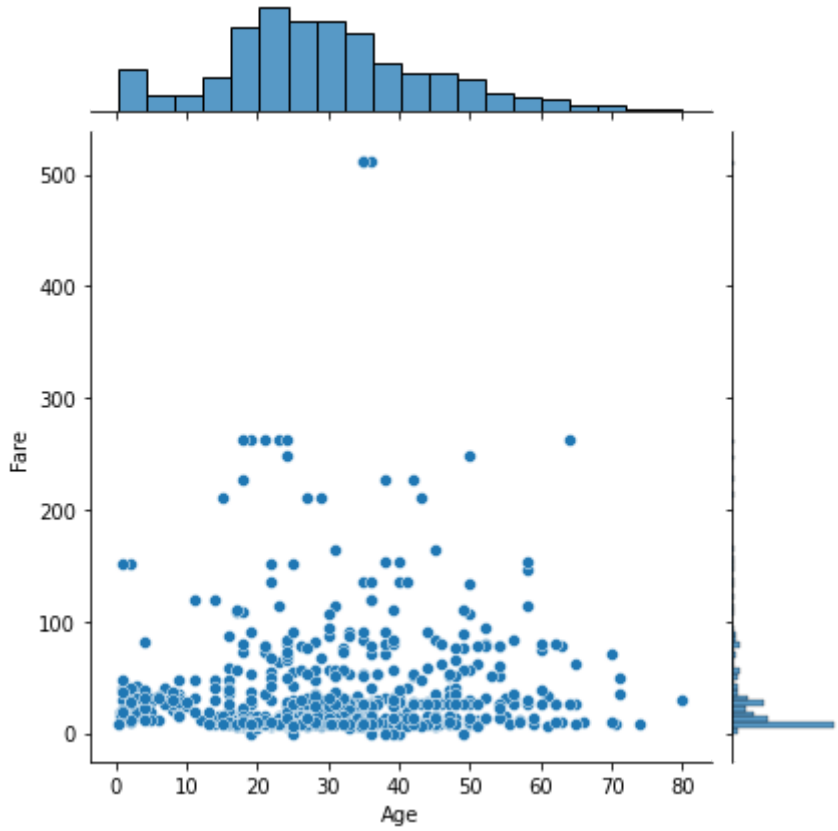
In [18]:
```
sns.jointplot(x = "Embarked", y = "Fare",
              kind = "scatter", data = df)
```

Out[18]:
```
<seaborn.axisgrid.JointGrid at 0x7f3e8daf9520>
```



In [19]:
```
sns.jointplot(x = "Age", y = "Fare",
              kind = "scatter", data = df)
```

Out[19]:
```
<seaborn.axisgrid.JointGrid at 0x7f3e8d8ce9d0>
```



In [ ]:
```

```