DSBDA Lab Assignment No. 8

```
Name: Akash Ganesh Padir
Roll No.: TEB04
```

In [42]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import math
import numpy as np
import seaborn as sns
df = sns.load_dataset('titanic')
```

In [43]:
```
df.head()
```

Out[43]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no

In [44]:
```
df.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
```

Use the Seaborn library to see if we can find any patterns in the data.

```
Patterns of data can be find out with the help of different types of plots Types of plots are:
1. Distribution Plots
 a. Dist-Plot
 b. Joint Plot
 c. pairplot
```

```
 d. Rug Plot
2. Categorical Plots
 a. Bar Plot
 b. Count Plot
 c. Box Plot
 d. Violin Plot
3. Advanced Plots
 a. Strip Plot
 b. Swarm Plot
4.Matrix Plots
 a. Heat Map
 b. Cluster Map
```

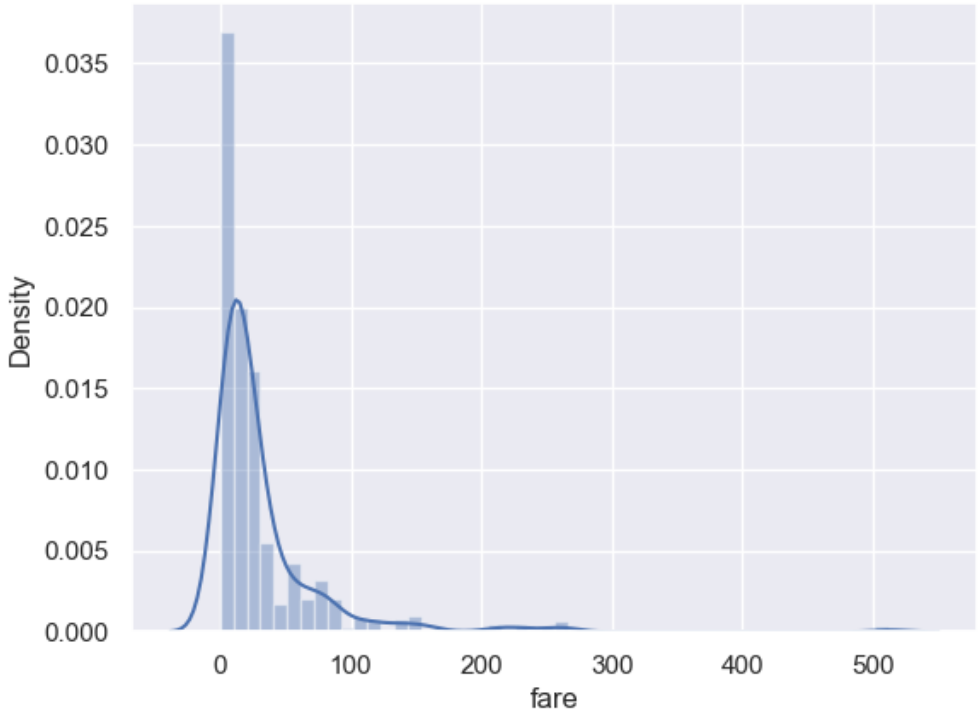
1. Distribution Plots:

```
a. Dist Plot
```

In [45]:
```
sns.distplot(df['fare'])
```

```
C:\Users\COMP 549\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `dis
tplot` 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 axe
s-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

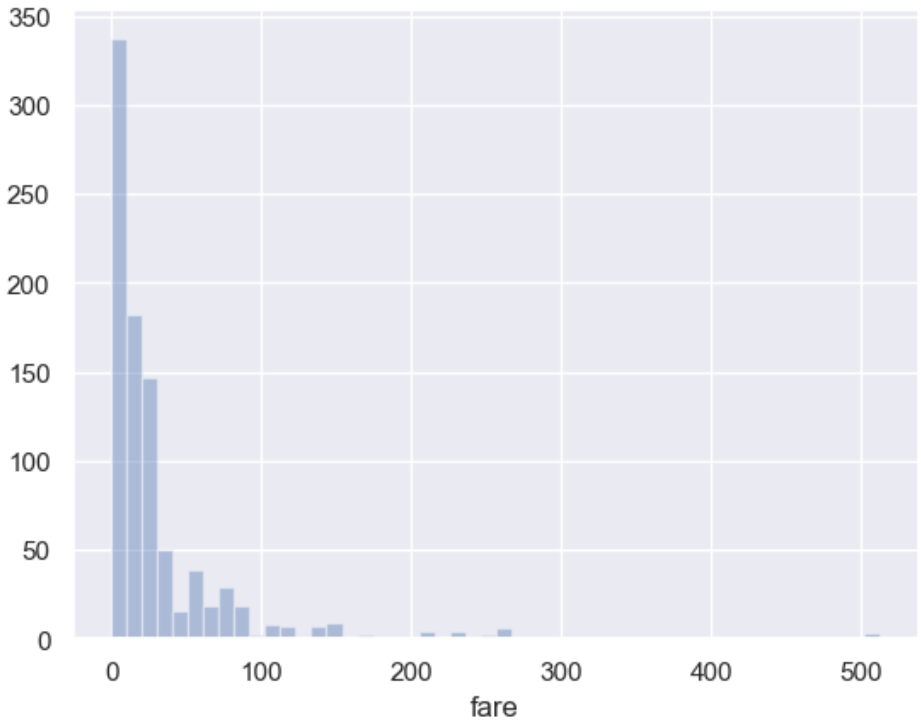
Out[45]: <AxesSubplot:xlabel='fare', ylabel='Density'>



In [46]:
```
sns.distplot(df["fare"],kde=False)
```

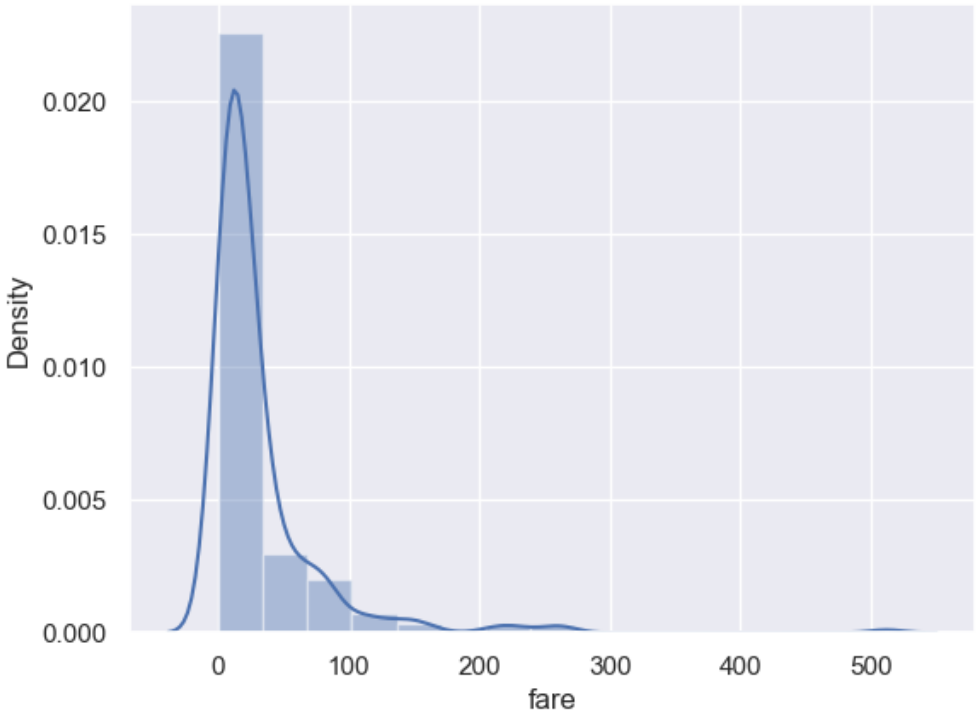
```
C:\Users\COMP 549\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `dis
tplot` is a deprecated function and will be removed in a future version. Please adapt your code
to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axe
s-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[46]: <AxesSubplot:xlabel='fare'>



In [47]:
```
sns.distplot(df["fare"],kde=True,bins=15)
```

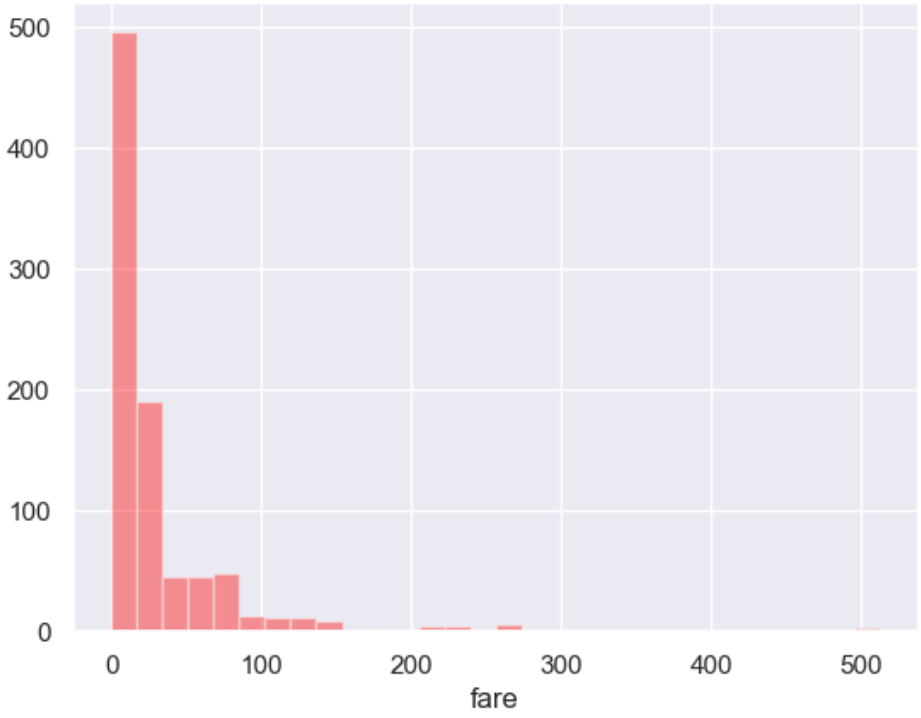
Out[47]: <AxesSubplot:xlabel='fare', ylabel='Density'>



In [48]:
```
sns.distplot(df["fare"],bins=30,kde=False,color='Red')
```

```
C:\Users\COMP 549\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `dis
tplot` 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 axe
s-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

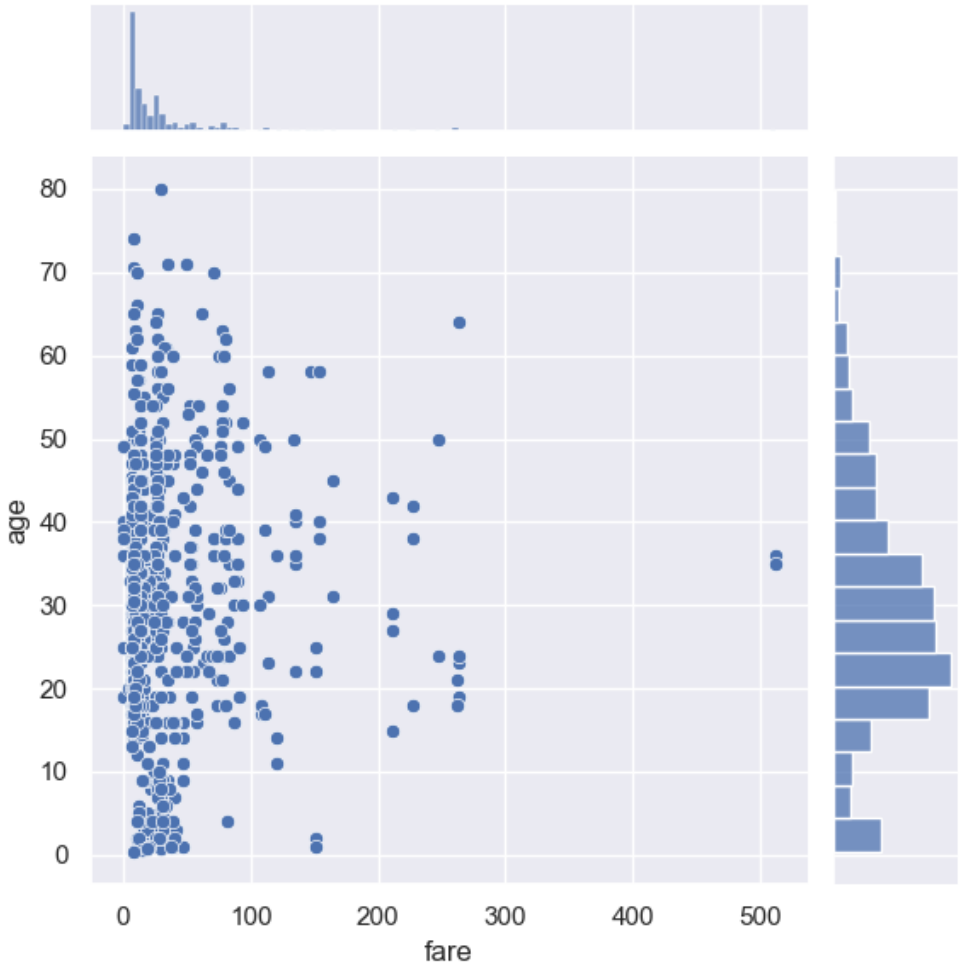
Out[48]: <AxesSubplot:xlabel='fare'>




```
B. Joint Plot
```

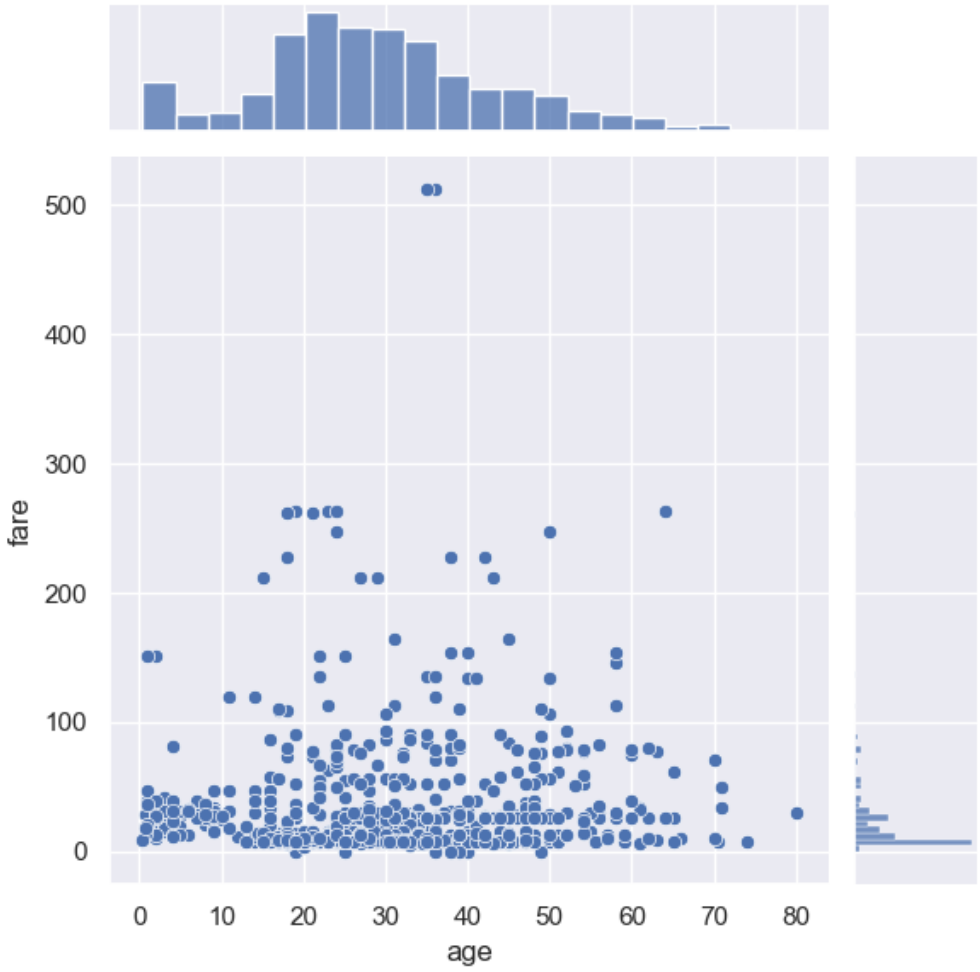
In [49]:
```
sns.jointplot(x='fare',y='age',data=df)
```

Out[49]:
```
<seaborn.axisgrid.JointGrid at 0x256b1a66b80>
```



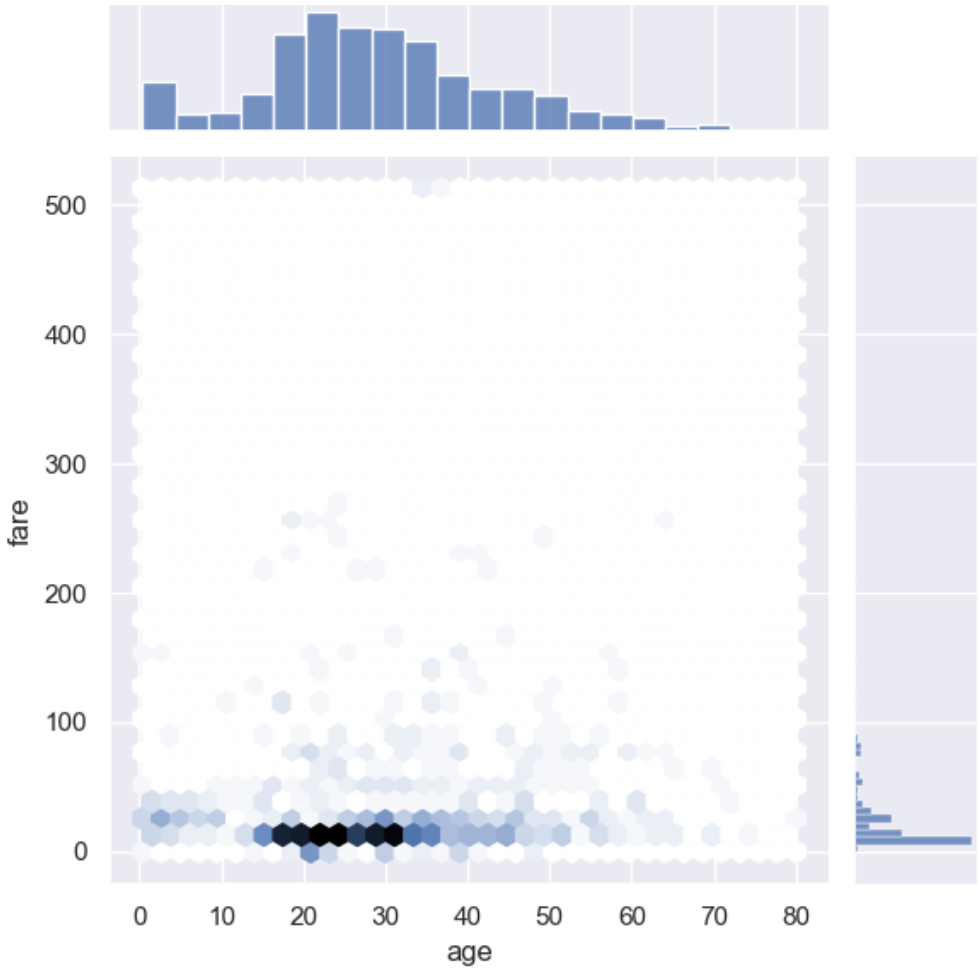
In [50]:
```
#For Plot 1
sns.jointplot(x = df['age'], y = df['fare'], kind = 'scatter')
```

Out[50]:
```
<seaborn.axisgrid.JointGrid at 0x256b4a22c70>
```



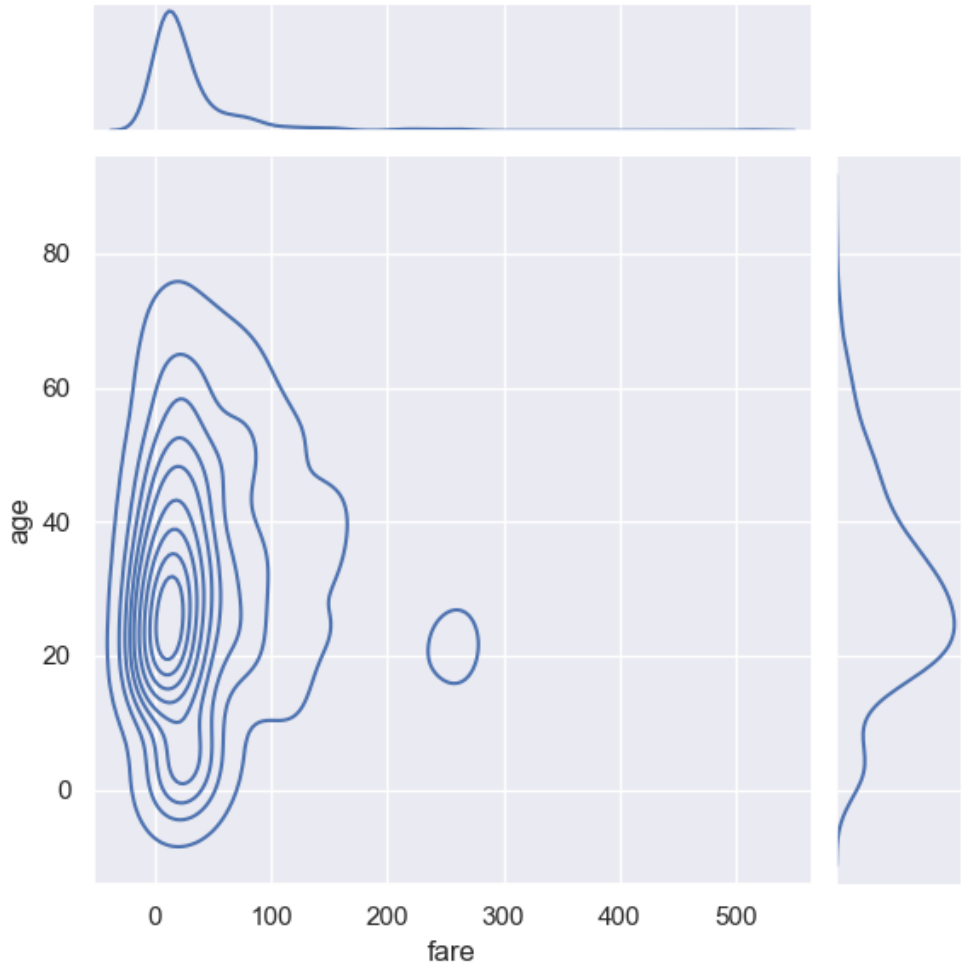
In [51]:
```
#For Plot 2
sns.jointplot(x = df['age'], y = df['fare'], kind = 'hex')
```

Out[51]: <seaborn.axisgrid.JointGrid at 0x256b1a66a30>



In [52]:
```
sns.jointplot(x='fare',y='age',data=df,kind='kde')
```

Out[52]:
```
<seaborn.axisgrid.JointGrid at 0x256b19dfa60>
```




```
C. Pair Plot
```

In [53]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
dataset=pd.read_csv("tested.csv")
dataset
```

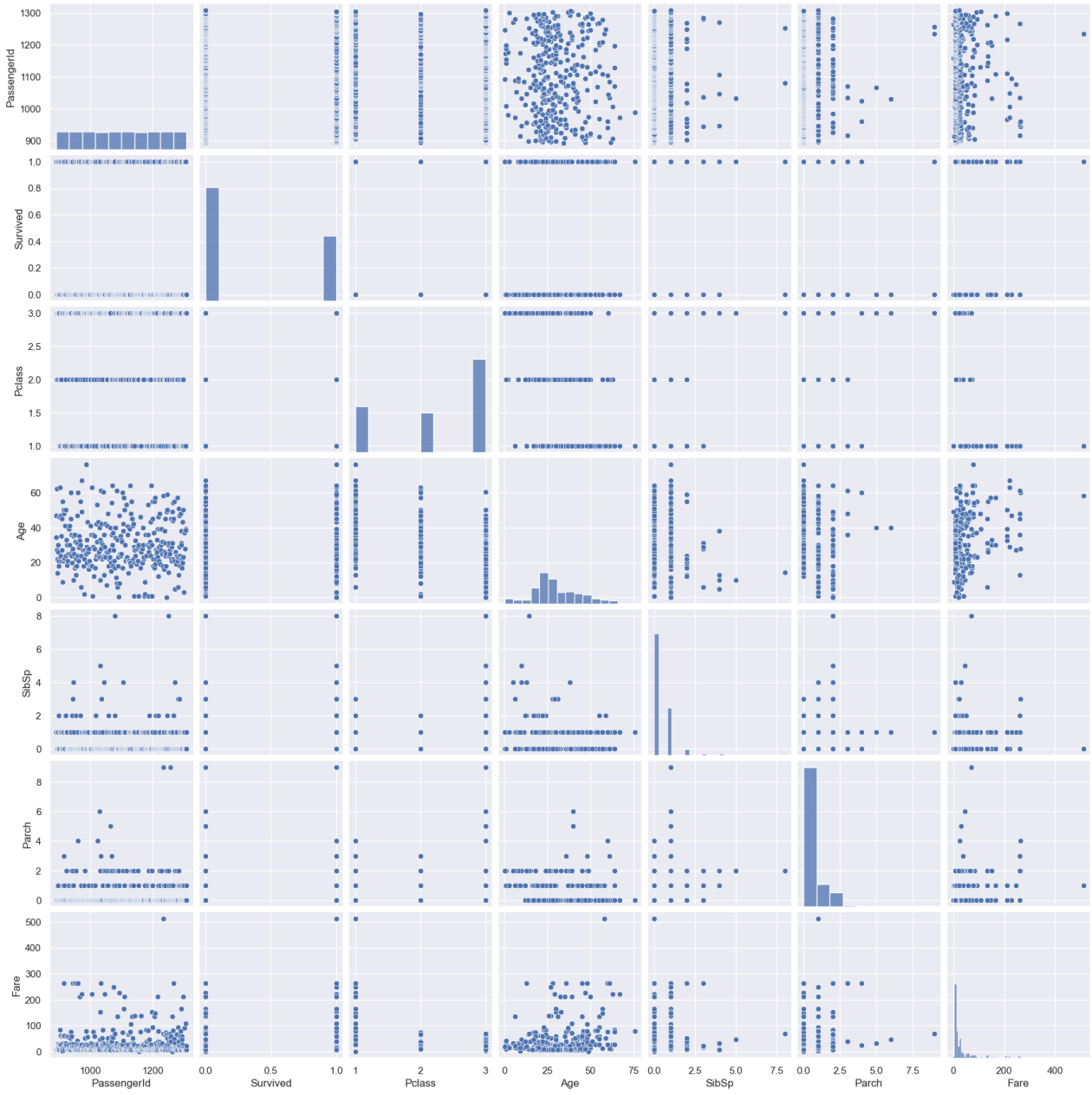
Out[53]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
413	1305	0	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	S
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	C
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	S
416	1308	0	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	S
417	1309	0	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	C

418 rows × 12 columns

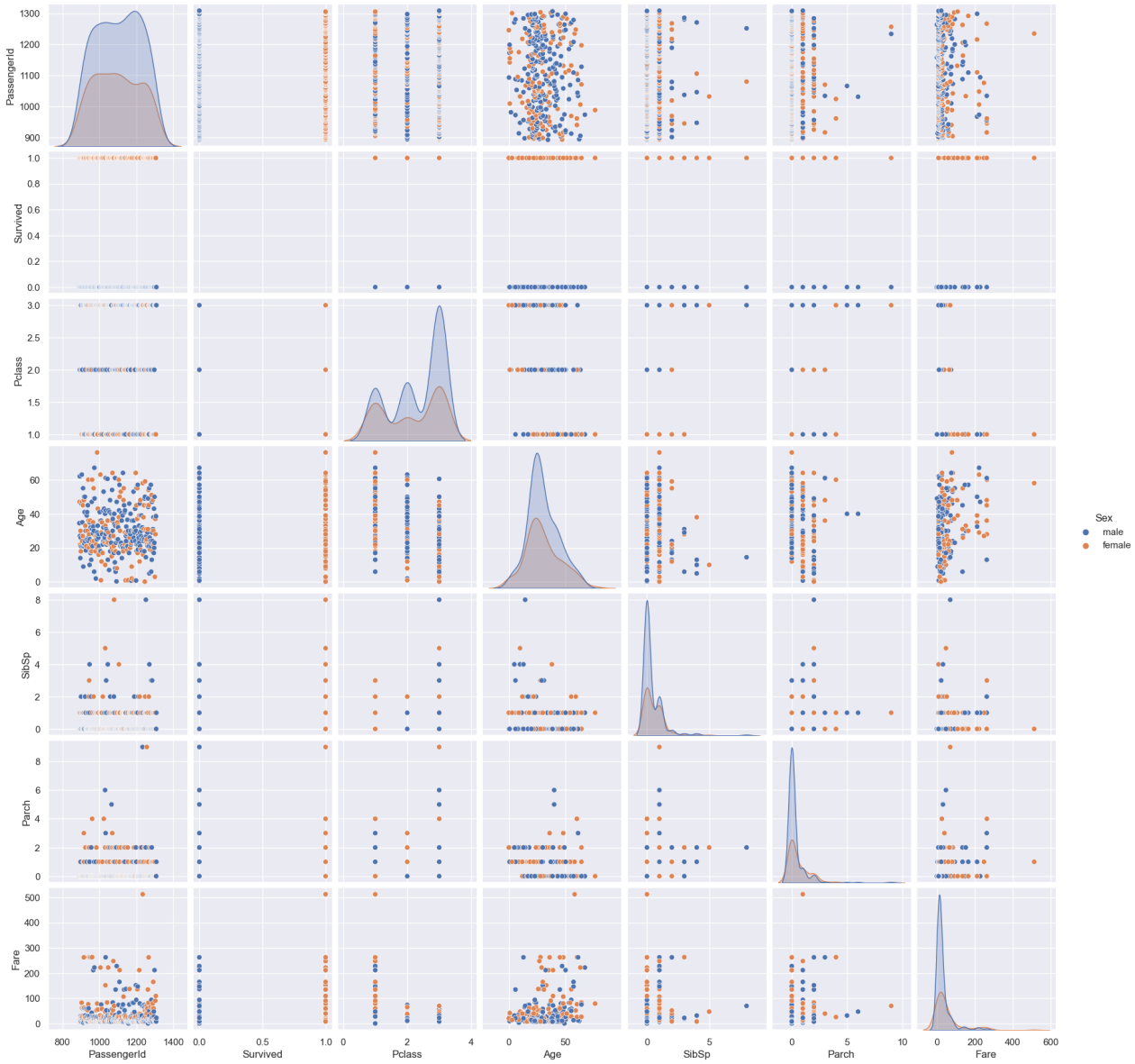
In [54]:
```
sns.pairplot(dataset)
```

Out[54]:
```
<seaborn.axisgrid.PairGrid at 0x256b71429a0>
```



In [55]:
```
sns.pairplot(dataset,hue="Sex")
```

Out[55]: <seaborn.axisgrid.PairGrid at 0x256b1a81760>



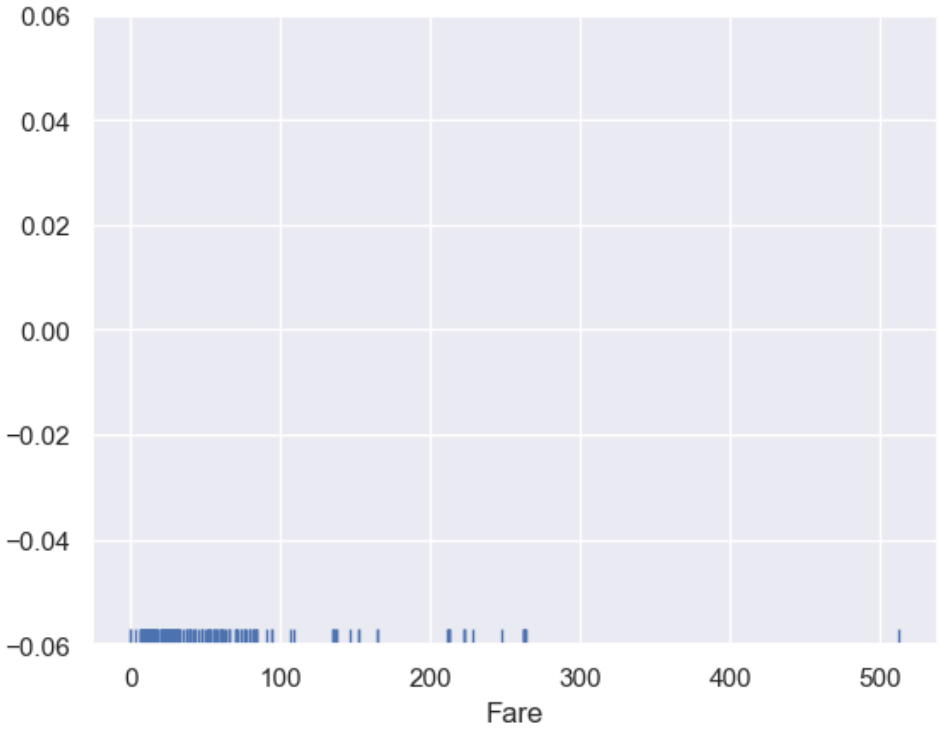
In [56]:
```
sns.pairplot(dataset, hue='Sex', diag_kind="hist", kind="scatter", palette="husl")
```

Out[56]:
```
<seaborn.axisgrid.PairGrid at 0x256b760cfd0>
```

D.Rug Plot

In [57]:
```
sns.rugplot(dataset['Fare'])
```

Out[57]:
```
<AxesSubplot:xlabel='Fare'>
```

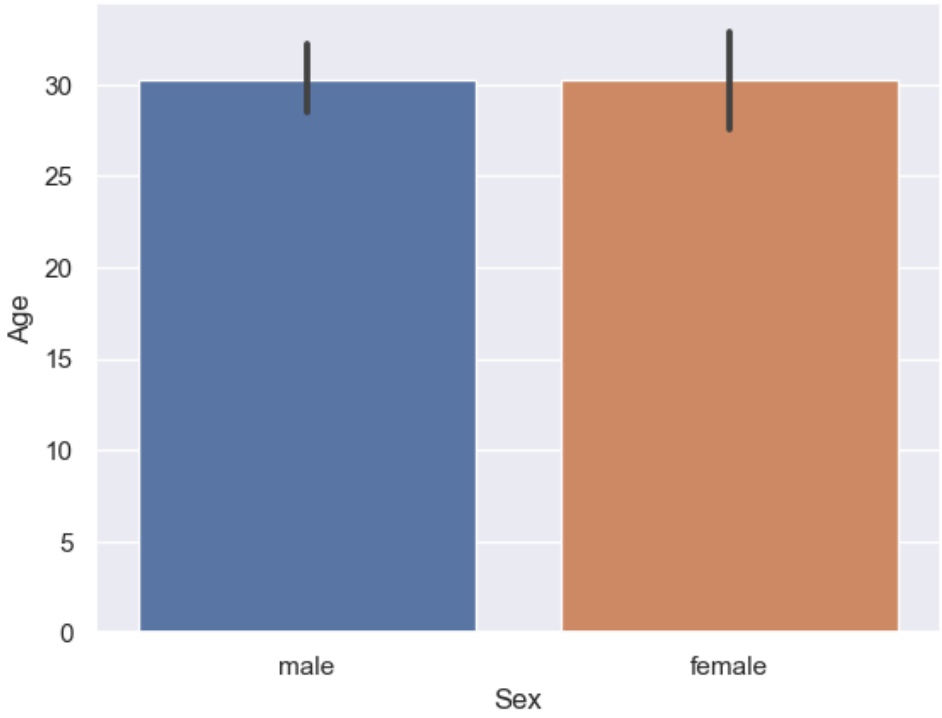


2. Categorical plots

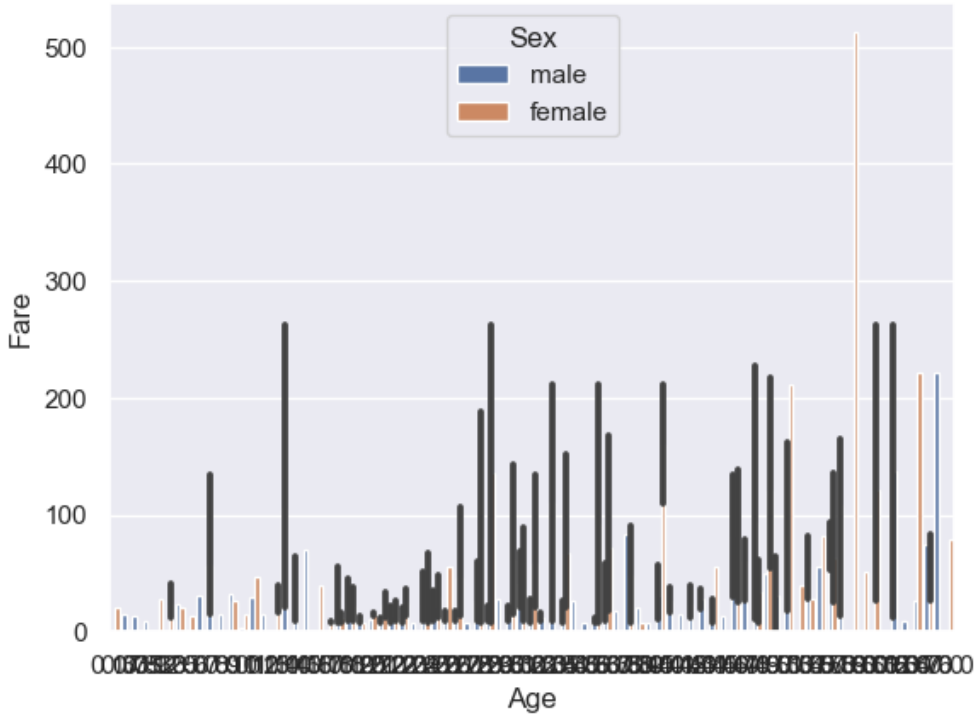
```
a.bar plot
```

In [58]:
```
sns.barplot(x='Sex', y='Age', data=dataset)
```

Out[58]: <AxesSubplot:xlabel='Sex', ylabel='Age'>



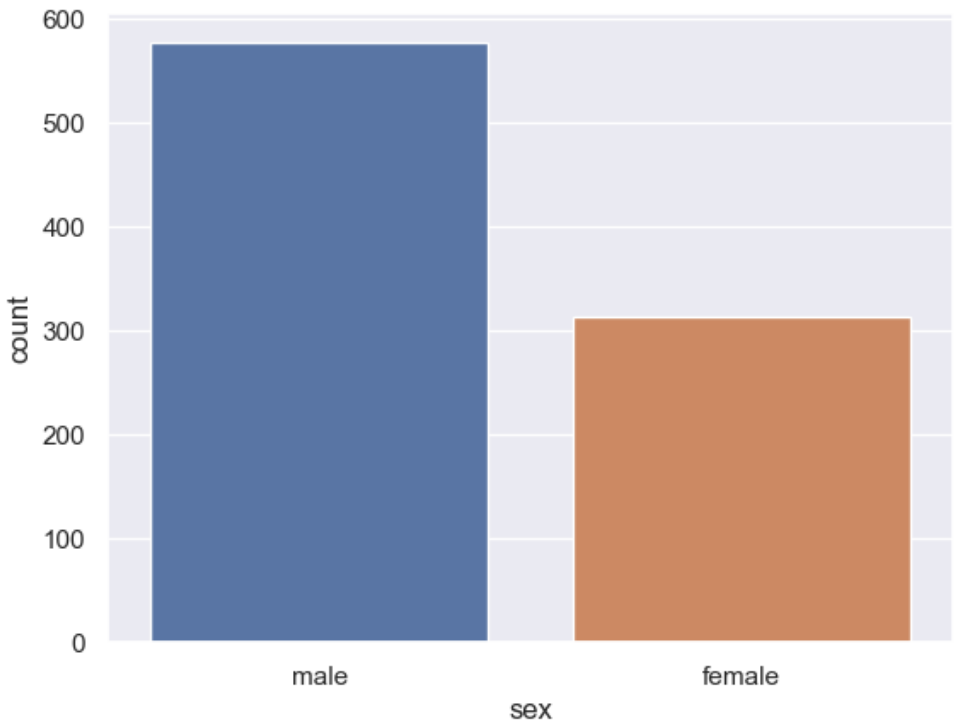
In [59]:
```
ax = sns.barplot(x="Age", y="Fare", hue="Sex", data=dataset)
```



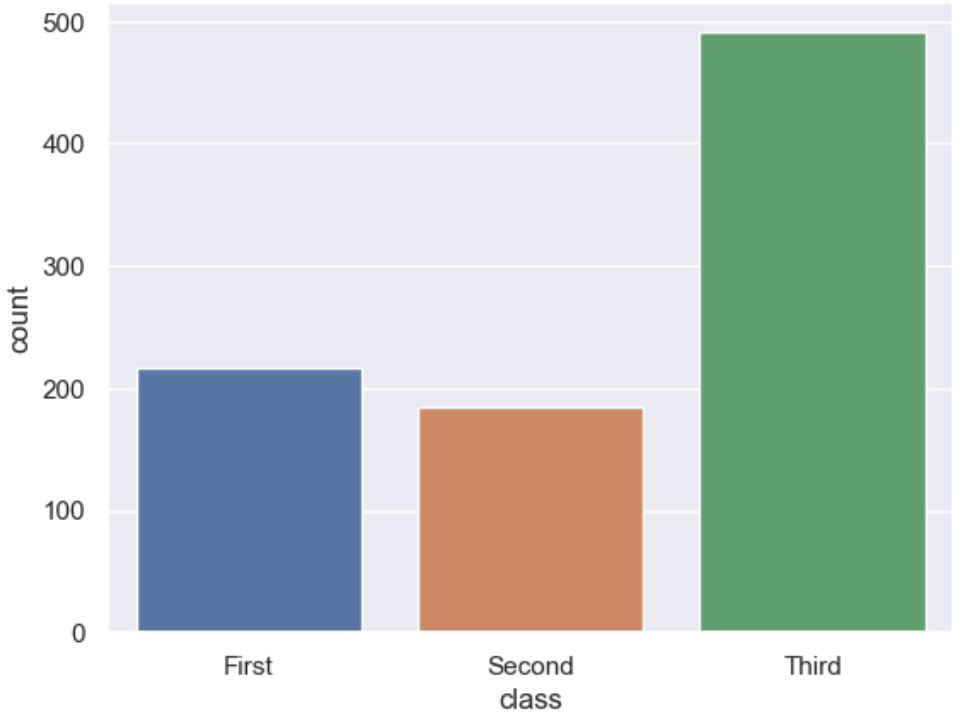

```
b. count plot
```

In [60]:
```
sns.countplot(x ='sex', data = df)
```

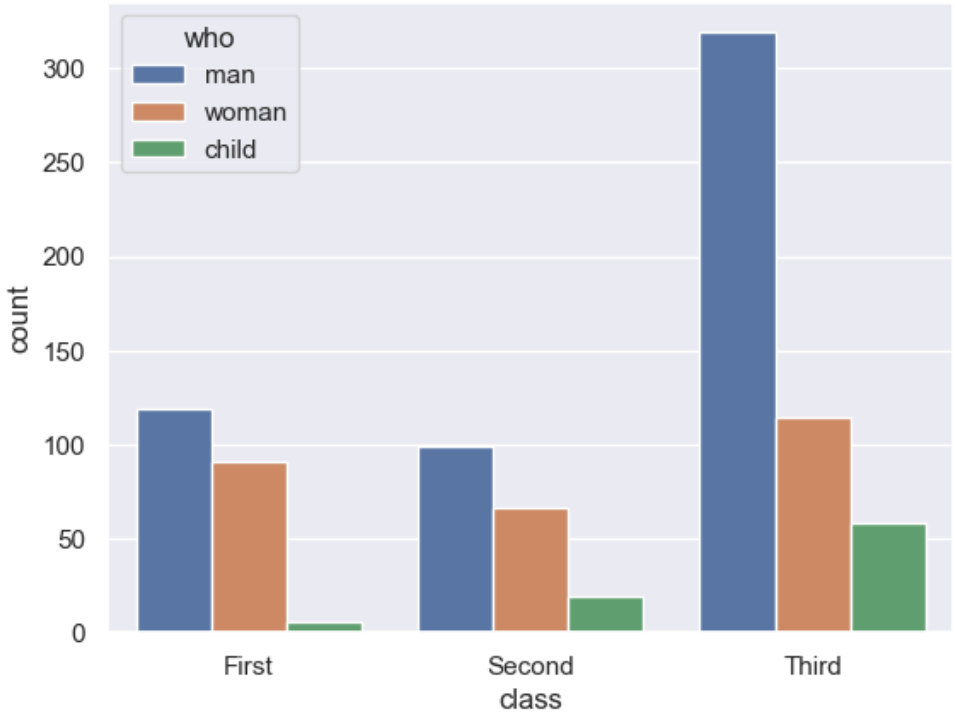
Out[60]:
```
<AxesSubplot:xlabel='sex', ylabel='count'>
```



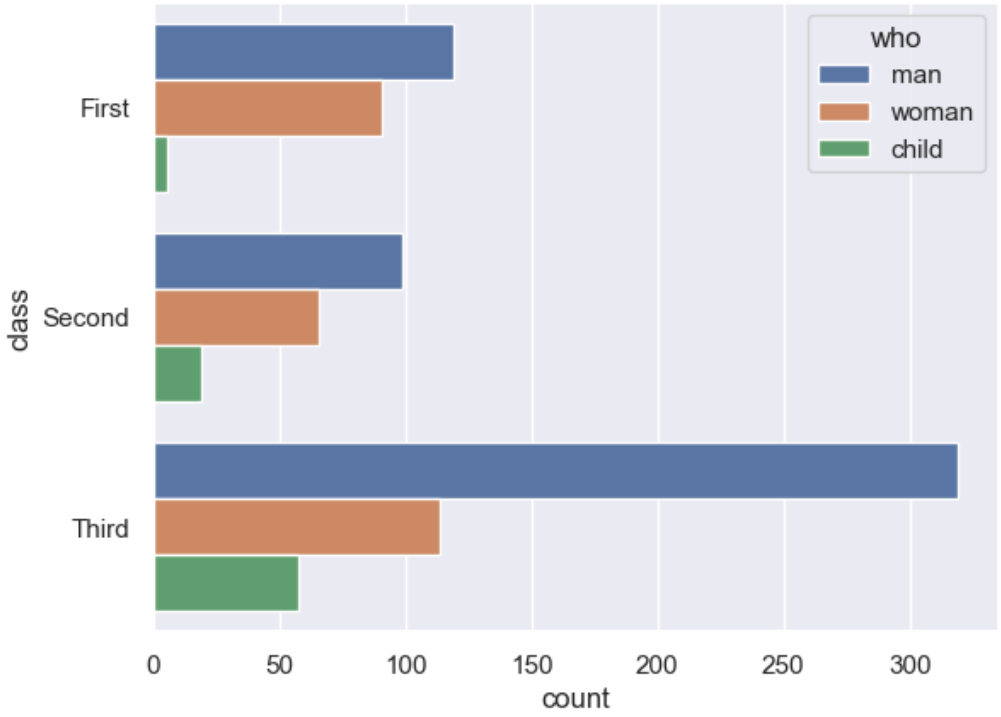
In [61]:
```
sns.set_theme(style="darkgrid")
ax = sns.countplot(x="class", data=df)
```



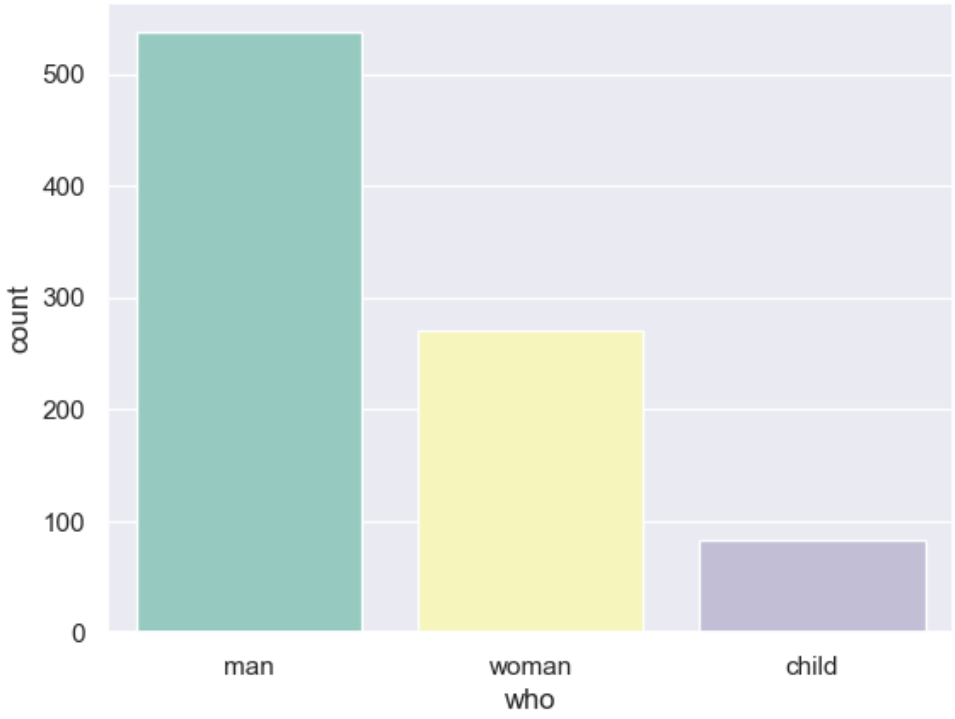
In [62]:
```
ax = sns.countplot(x="class", hue="who", data=df)
```



In [63]:
```
ax = sns.countplot(y="class", hue="who", data=df)
```

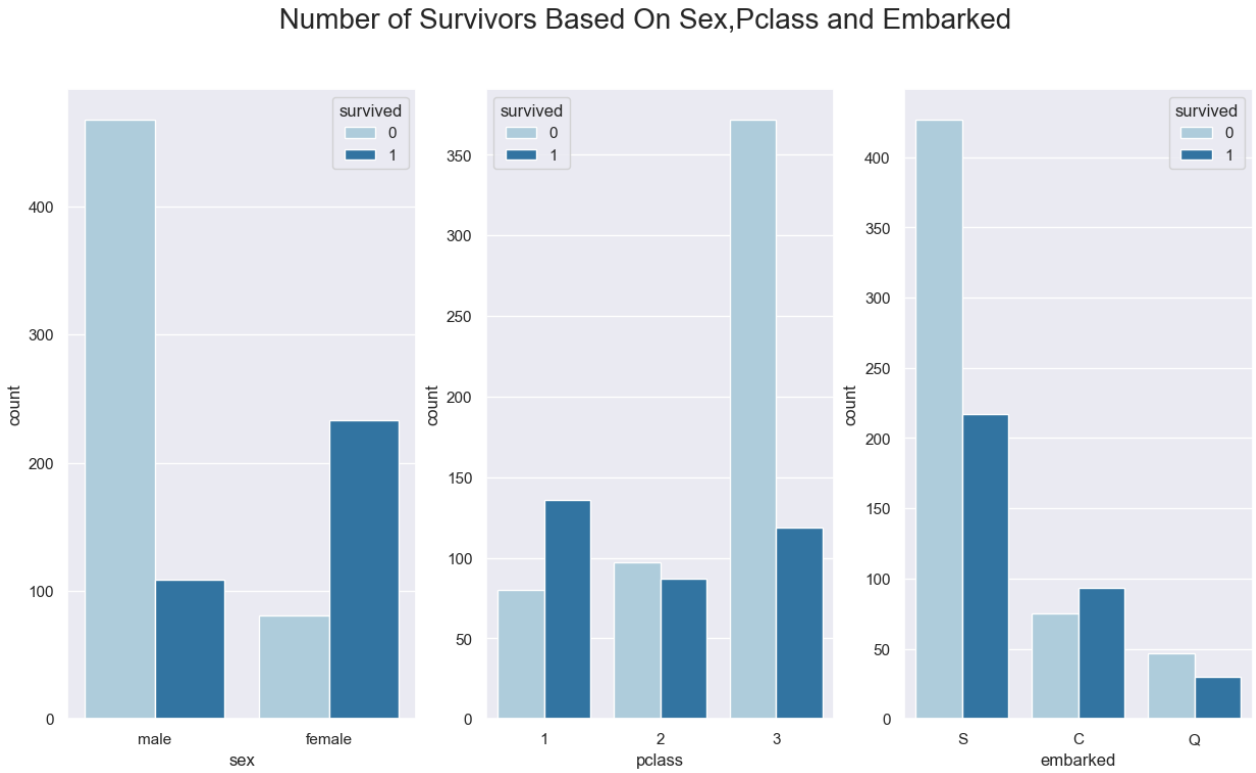


In [29]:
```
ax = sns.countplot(x="who", data=df, palette="Set3")
```



In [30]:
```
fig,axes=plt.subplots(1,3,figsize=(15,8))
plt.suptitle(" Number of Survivors Based On Sex,Pclass and Embarked",fontsize=20)
sns.countplot(x="sex",hue="survived",data=df,ax=axes[0],palette="Paired")
sns.countplot(x="pclass",hue="survived",data=df,ax=axes[1],palette="Paired")
sns.countplot(x="embarked",hue="survived",data=df,ax=axes[2],palette="Paired")
```

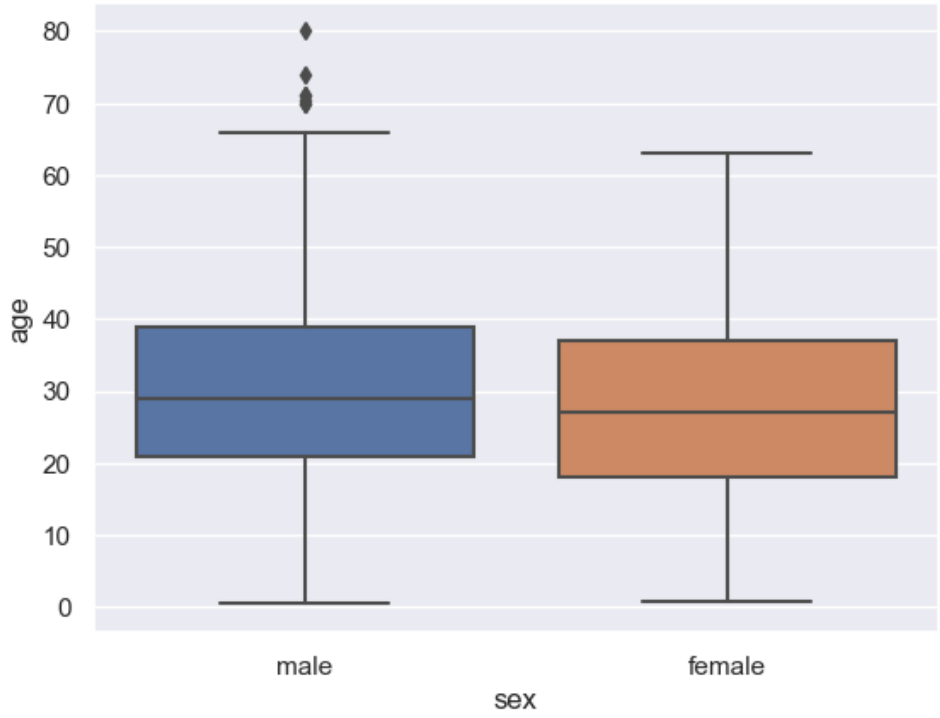
Out[30]: <AxesSubplot:xlabel='embarked', ylabel='count'>



C. Box Plot

In [33]:
```
sns.boxplot(x='sex', y='age', data=df)
```

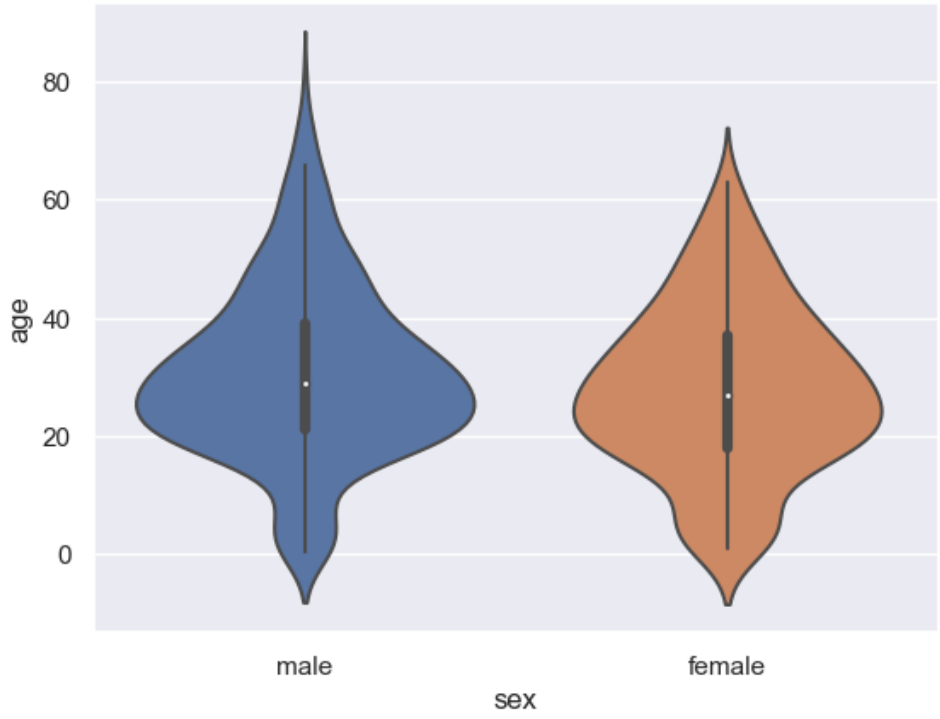
Out[33]: <AxesSubplot:xlabel='sex', ylabel='age'>



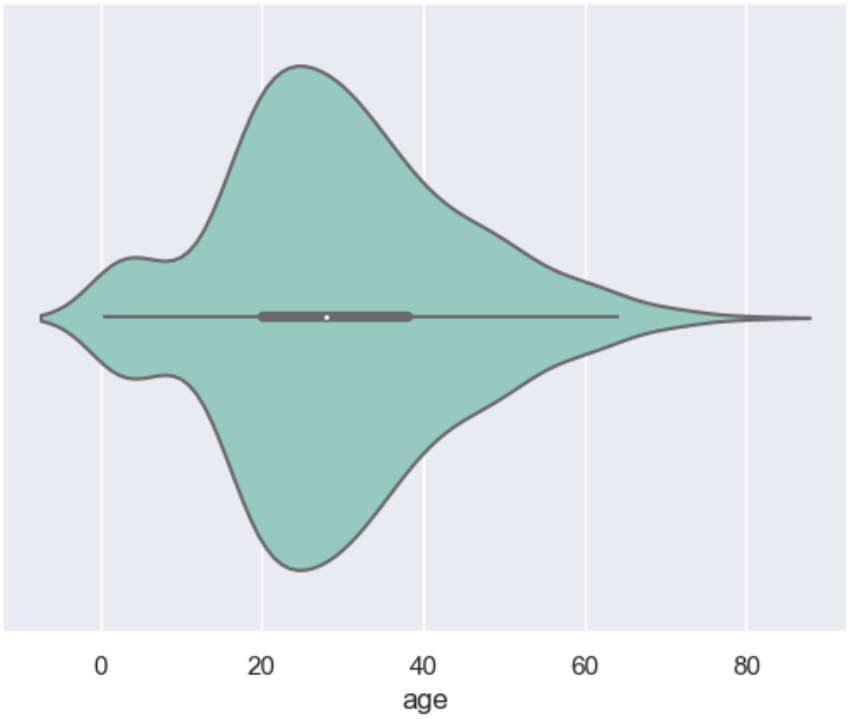
d. Violin Plot

In [34]:
```
sns.violinplot(x='sex', y='age', data=df)
```

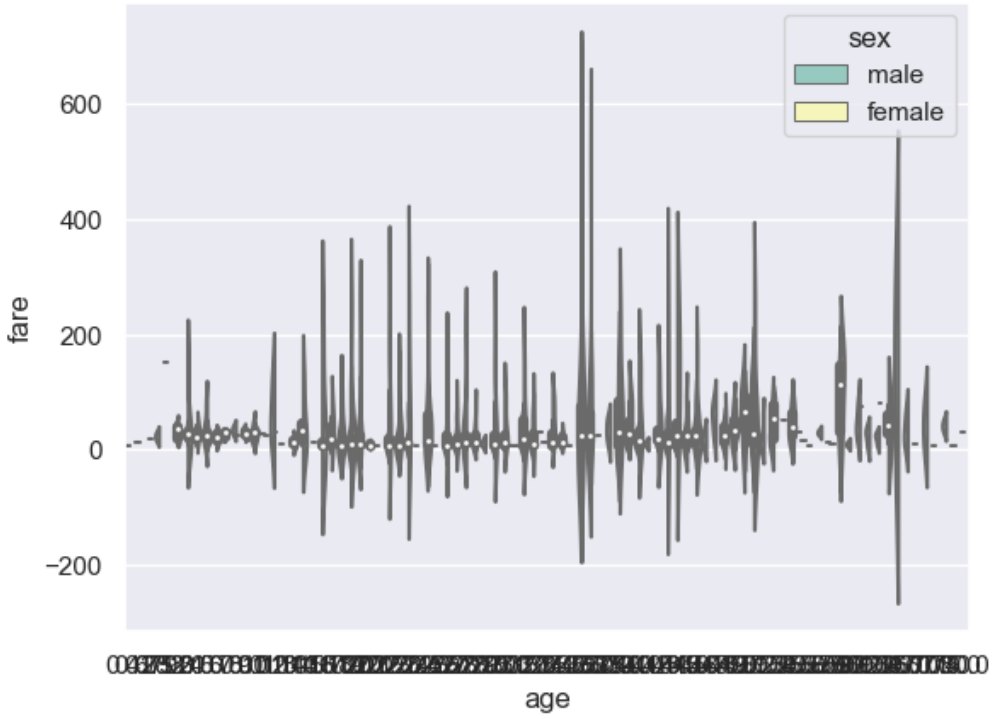
Out[34]: <AxesSubplot:xlabel='sex', ylabel='age'>



In [35]:
```
ax = sns.violinplot(x="age", hue="sex",data=df, palette="Set3", split=True,scale="count")
```

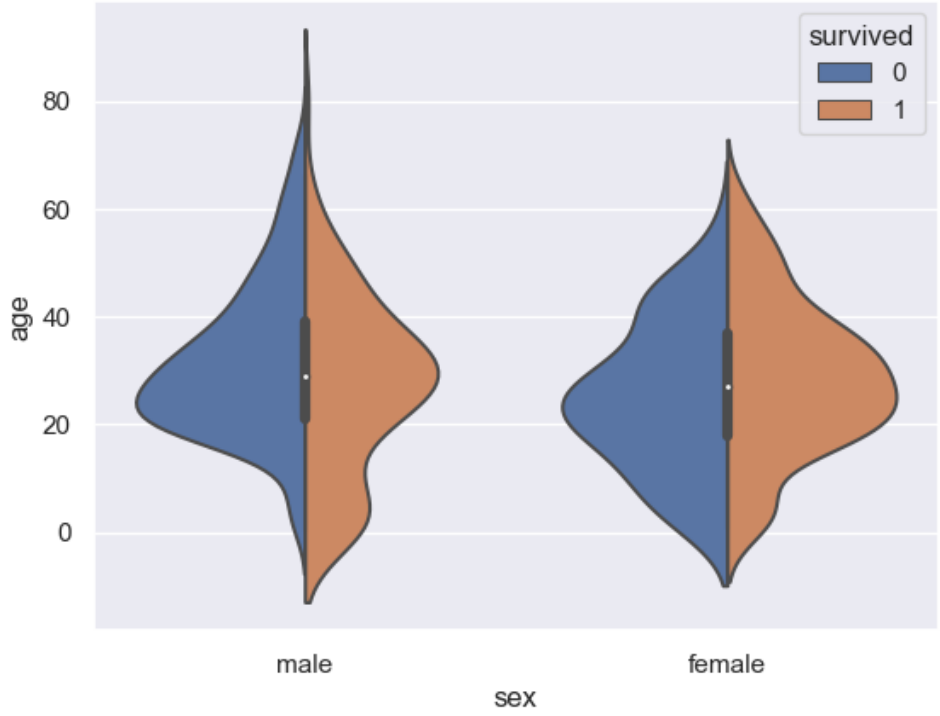


In [36]:
```
ax = sns.violinplot(x="age",y="fare", hue="sex",data=df, palette="Set3", split=True,scale="count"
```



In [37]:
```
sns.violinplot(x ="sex", y ="age", hue ="survived",data = df, split = True)
```

Out[37]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```

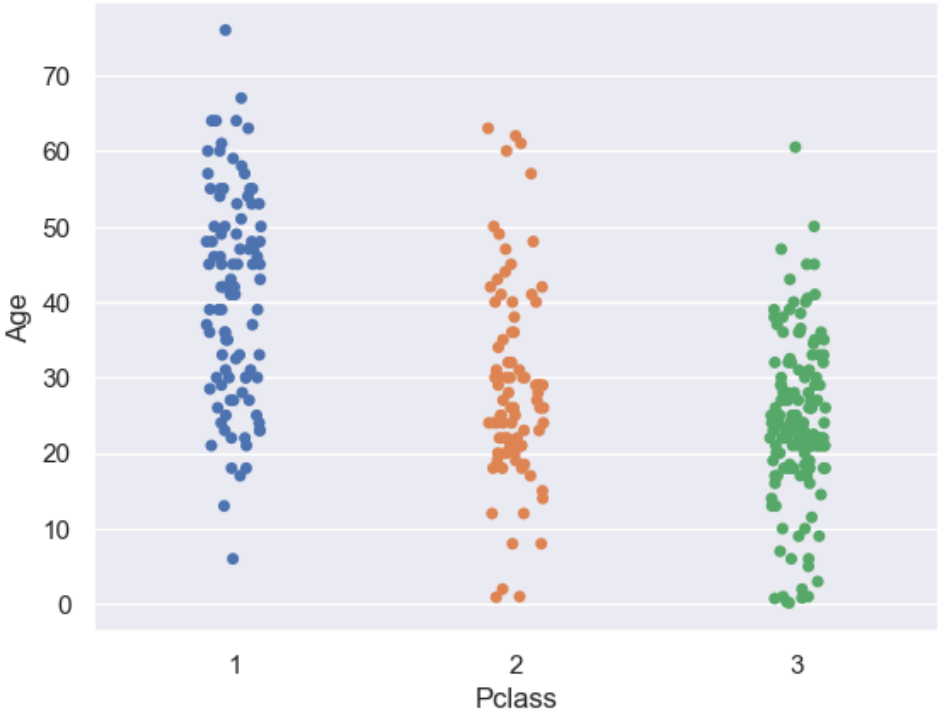


3. Advanced Plots

```
a. Strip Plot
```

In [38]:
```
sns.stripplot(y = dataset['Age'], x = dataset['Pclass'])
```

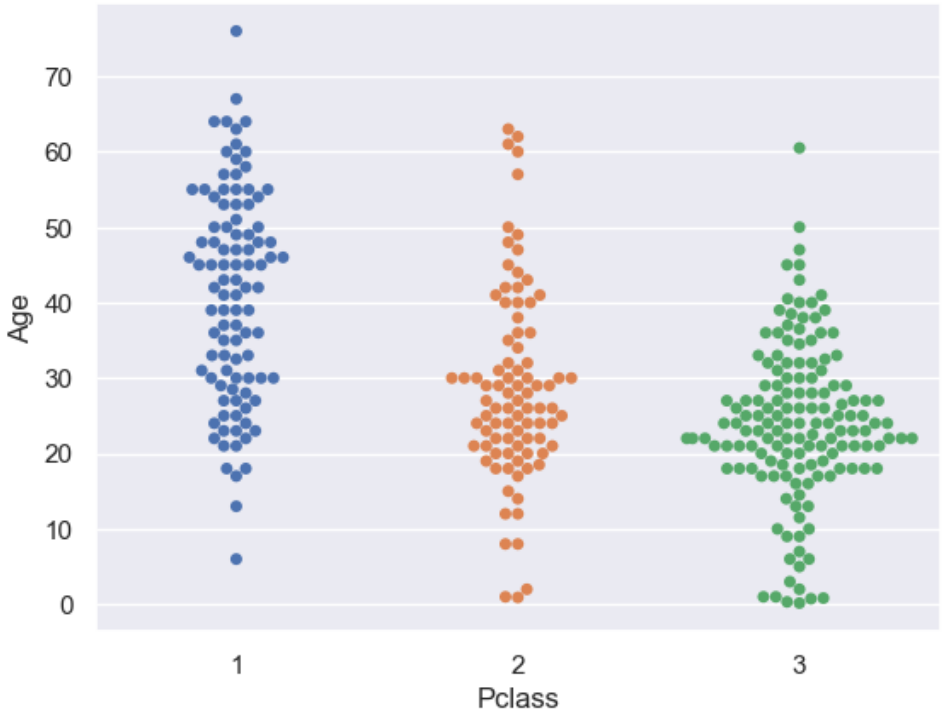
Out[38]:
```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



```
b. Swarm Plot
```

In [39]:
```
sns.swarmplot(y = dataset['Age'], x = dataset['Pclass'])
```

Out[39]:
```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```

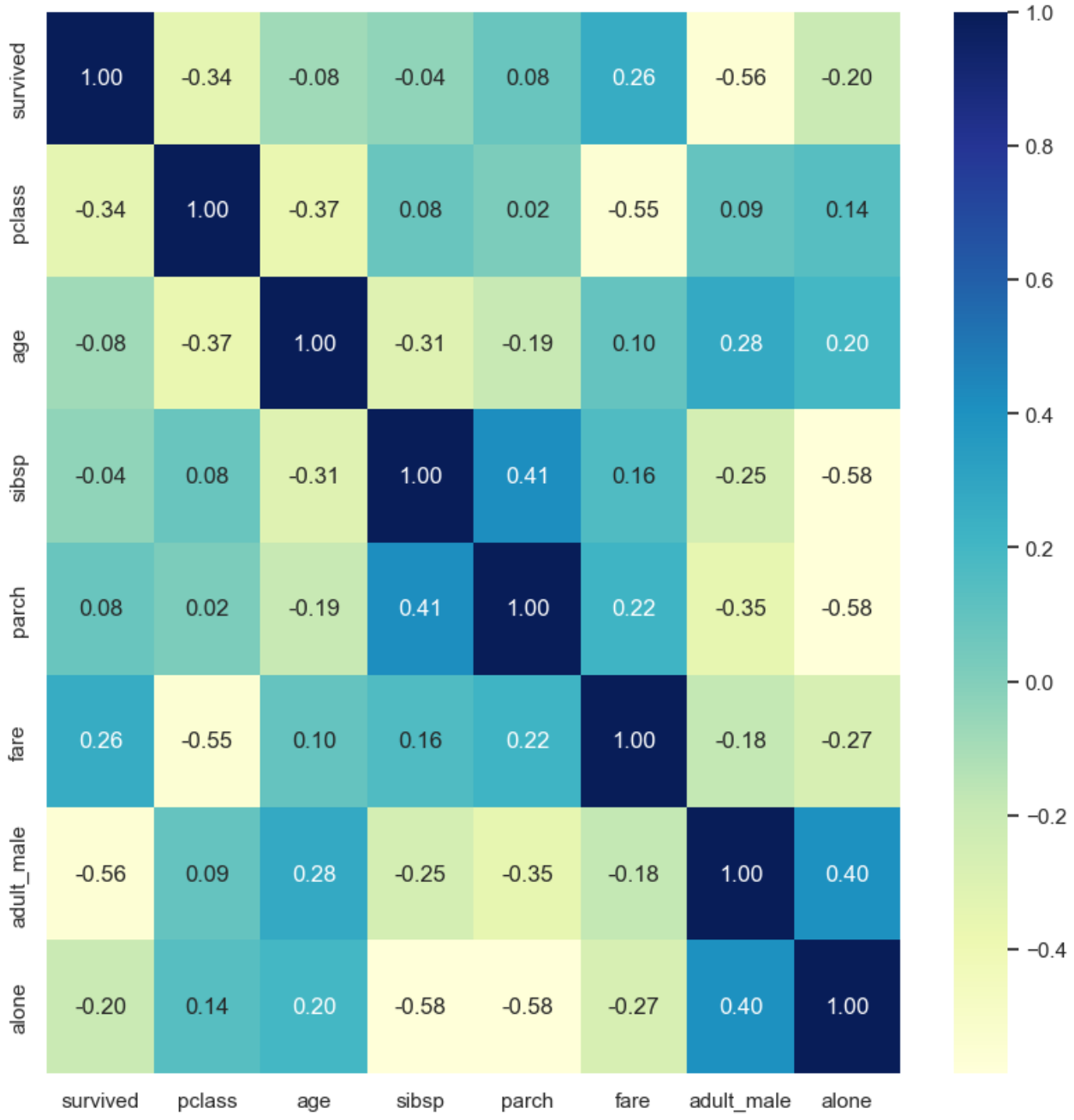


D. Matrix plots

```
a.heatmap
```

In [40]:
```
plt.subplots(figsize=(10, 10))
sns.heatmap(df.corr(), cmap = "YlGnBu", annot=True, fmt=".2f")
```

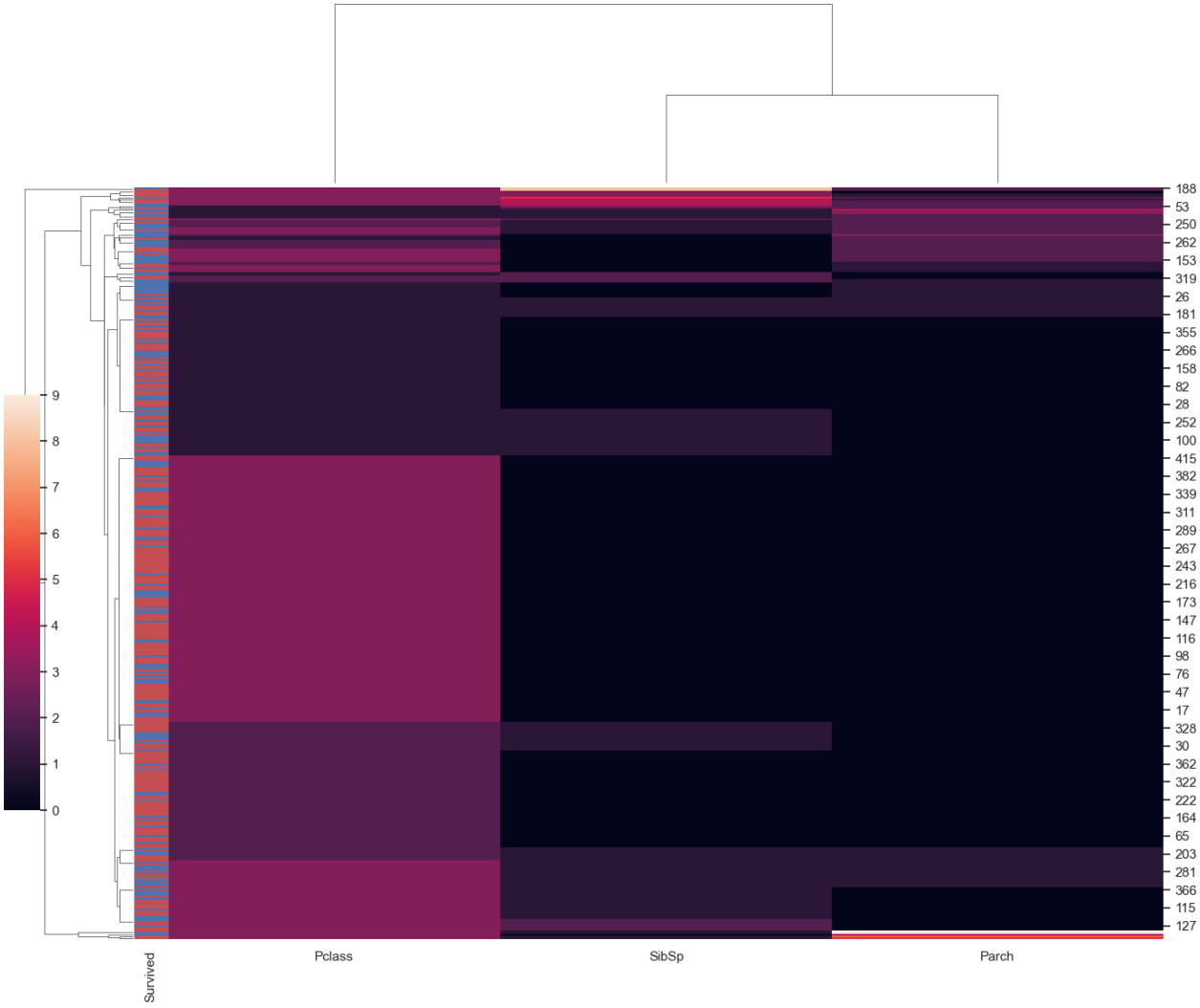
Out[40]: <AxesSubplot:>




```
b.clustermap
```

In [41]:
```
data = dataset[["Pclass", "SibSp", "Parch"]]
survived = dataset["Survived"]
lut = dict(zip(survived.unique(), "rb"))
row_colors = survived.map(lut)
sns.clustermap(data, figsize=(14,12),
 row_colors=row_colors,
 dendrogram_ratio=(.1, .2),
 cbar_pos=(0, .2, .03, .4))
```

Out[41]: <seaborn.matrix.ClusterGrid at 0x256b34a8bb0>



In [ ]: