In [1]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

dataset = sns.load_dataset('titanic')

dataset.head(10)
```

Out[1]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	d
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	N
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	N
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	N
5	0	3	male	NaN	0	0	8.4583	Q	Third	man	True	N
6	0	1	male	54.0	0	0	51.8625	S	First	man	True	
7	0	3	male	2.0	3	1	21.0750	S	Third	child	False	N
8	1	3	female	27.0	0	2	11.1333	S	Third	woman	False	N
9	1	2	female	14.0	1	0	30.0708	C	Second	child	False	N

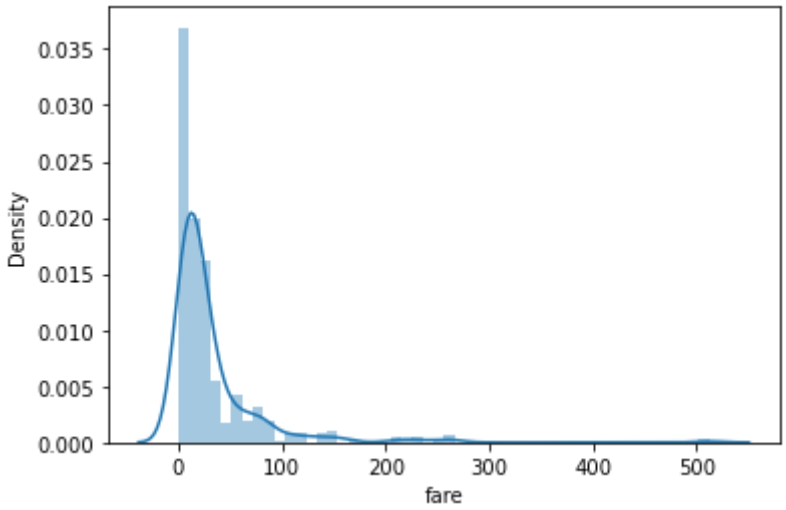
In [133…
```
dataset.info()adult_maleadult_male
```

```
<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 [134…
```
sns.distplot(dataset['fare'])
```

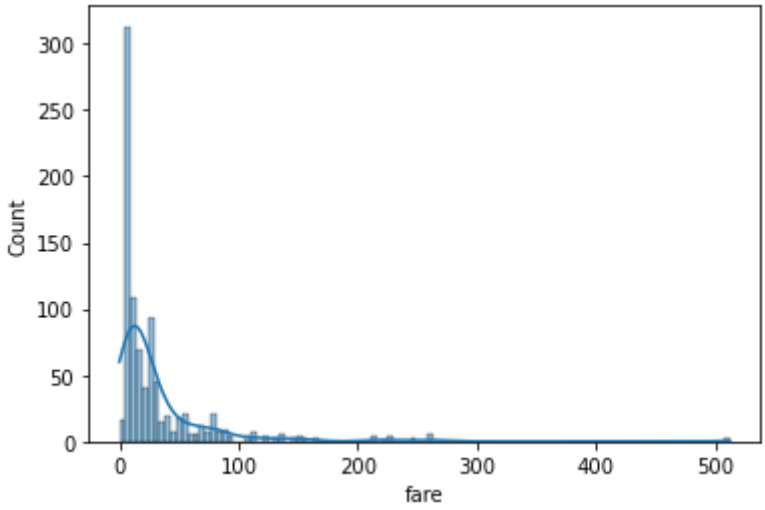
```
/home/pktc-320/anaconda3/lib/python3.9/site-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[134…
```
<AxesSubplot:xlabel='fare', ylabel='Density'>
```



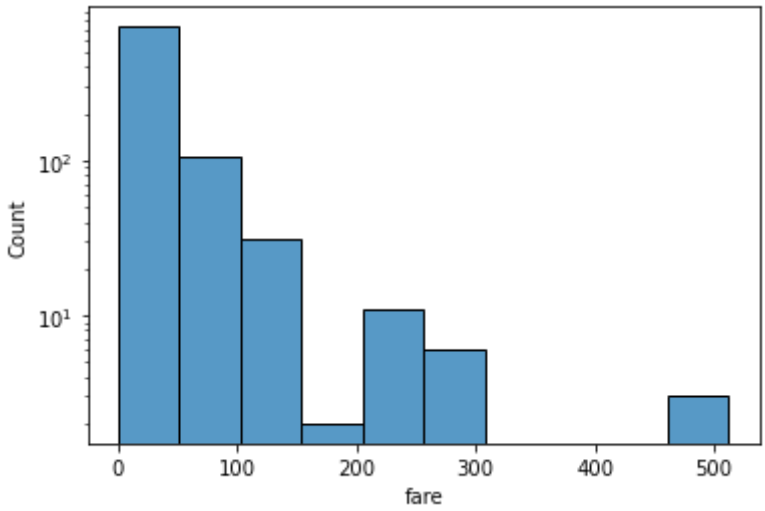
In [135…
```
#@sns.distplot(dataset['fare'], kde=False)
#sns.histplot(dataset['fare'], kde=False)
sns.histplot(dataset['fare'], kde=True)
```

Out[135… <AxesSubplot:xlabel='fare', ylabel='Count'>



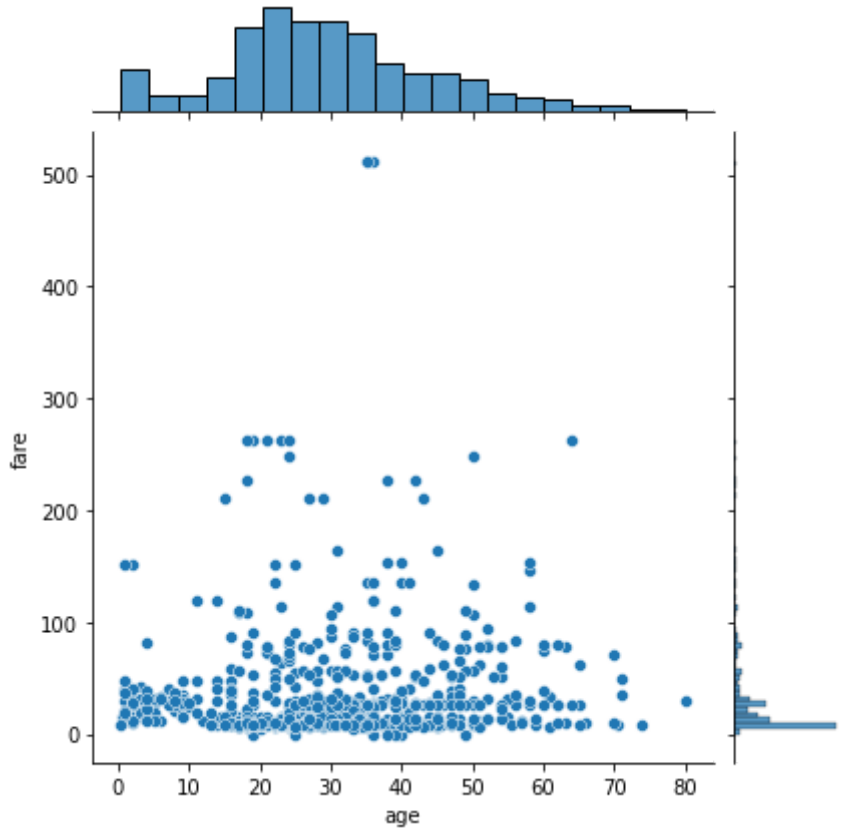
In [136…
```
#sns.distplot(dataset['fare'], kde=False, bins=10)
sns.histplot(dataset['fare'], kde=False, bins=10, log_scale=(0,10))
```

Out[136… <AxesSubplot:xlabel='fare', ylabel='Count'>



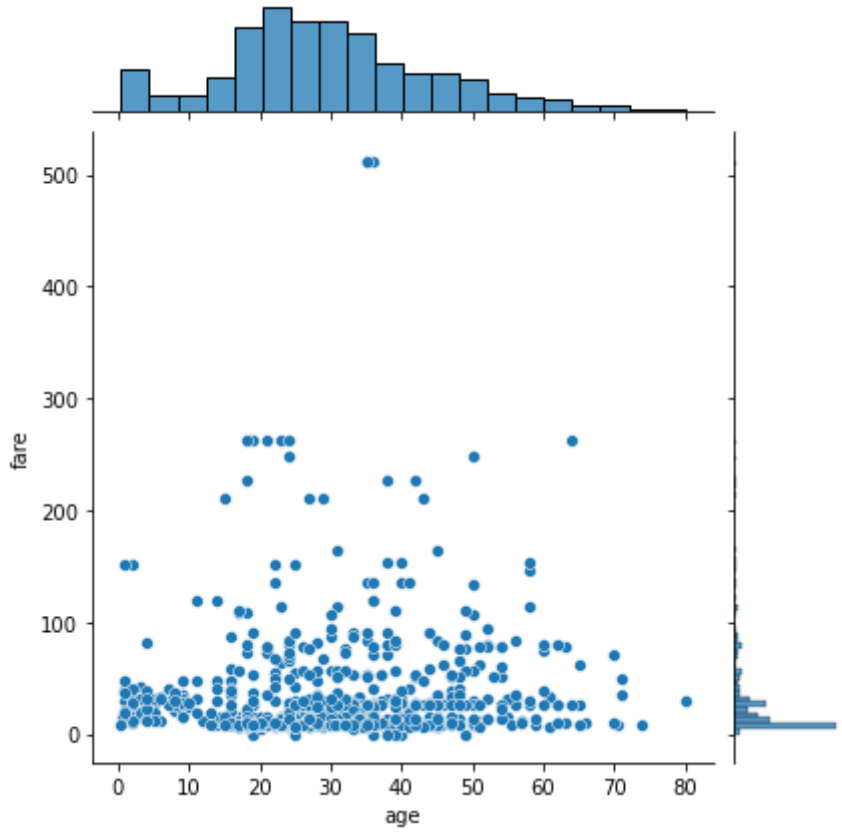
In [137…
```
sns.jointplot(x='age', y='fare', data=dataset)
```

Out[137… <seaborn.axisgrid.JointGrid at 0x7fbed56bb130>



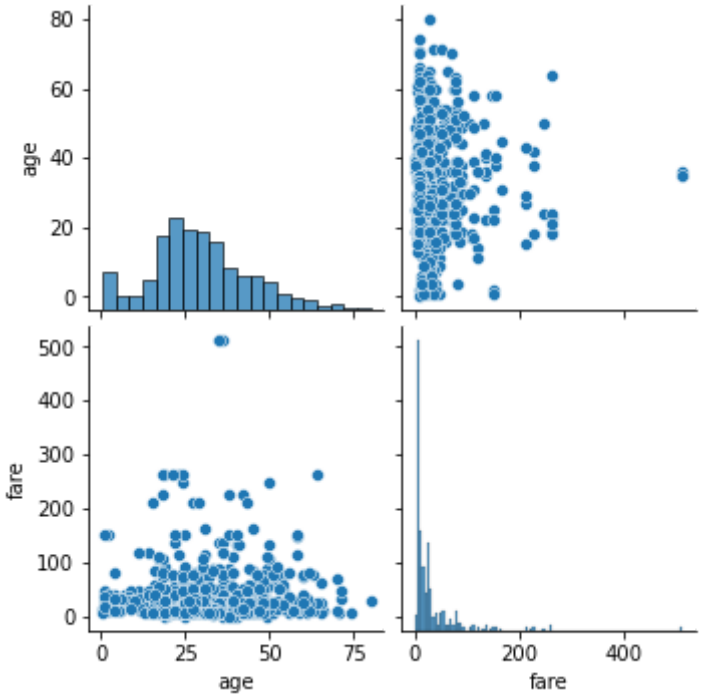
In [138…
```
#sns.jointplot(x='age', y='fare', data=dataset, kind='reg')
#sns.jointplot(x='age', y='fare', data=dataset, kind='hex')
sns.jointplot(x='age', y='fare', data=dataset, kind='scatter')
```

Out[138… <seaborn.axisgrid.JointGrid at 0x7fbed5f16190>



In [139…
```
sns.pairplot(dataset[['age','fare']])
```

Out[139… <seaborn.axisgrid.PairGrid at 0x7fbed5dff1c0>



In [140…
```
dataset.isna()
```

Out[140…

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	e
0	False	False	False	False	False	False	False	False	False	False	False	True	
1	False	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	True	
3	False	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	True	
…	…	…	…	…	…	…	…	…	…	…	…	…	
886	False	False	False	False	False	False	False	False	False	False	False	True	
887	False	False	False	False	False	False	False	False	False	False	False	False	
888	False	False	False	True	False	False	False	False	False	False	False	True	
889	False	False	False	False	False	False	False	False	False	False	False	False	
890	False	False	False	False	False	False	False	False	False	False	False	True	

891 rows × 15 columns

In [141…
```
#dataset['deck'].isnull()
dataset[dataset['deck'].isnull()]
```

Out[141…

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True
5	0	3	male	NaN	0	0	8.4583	Q	Third	man	True

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
7	0	3	male	2.0	3	1	21.0750	S	Third	child	False
...	...	...	...	...	...	...	...	...	...	...	...
884	0	3	male	25.0	0	0	7.0500	S	Third	man	True
885	0	3	female	39.0	0	5	29.1250	Q	Third	woman	False
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True

688 rows × 15 columns

In [142…
```
dataset[dataset['age'].isnull()]
```

Out[142…

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
5	0	3	male	NaN	0	0	8.4583	Q	Third	man	True
17	1	2	male	NaN	0	0	13.0000	S	Second	man	True
19	1	3	female	NaN	0	0	7.2250	C	Third	woman	False
26	0	3	male	NaN	0	0	7.2250	C	Third	man	True
28	1	3	female	NaN	0	0	7.8792	Q	Third	woman	False
...	...	...	...	...	...	...	...	...	...	...	...
859	0	3	male	NaN	0	0	7.2292	C	Third	man	True
863	0	3	female	NaN	8	2	69.5500	S	Third	woman	False
868	0	3	male	NaN	0	0	9.5000	S	Third	man	True
878	0	3	male	NaN	0	0	7.8958	S	Third	man	True
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False

177 rows × 15 columns

In [143…
```
dataset.dropna()
```

Out[143…

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	d
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	
6	0	1	male	54.0	0	0	51.8625	S	First	man	True	
10	1	3	female	4.0	1	1	16.7000	S	Third	child	False	
11	1	1	female	58.0	0	0	26.5500	S	First	woman	False	
...	...	...	...	...	...	...	...	...	...	...	...	
871	1	1	female	47.0	1	1	52.5542	S	First	woman	False	
872	0	1	male	33.0	0	0	5.0000	S	First	man	True	

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	d
879	1	1	female	56.0	0	1	83.1583	C	First	woman	False	
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	

182 rows × 15 columns

In [157…
```
dataset['sex'].value_counts()
```

Out[157…
```
male      577
female    314
Name: sex, dtype: int64
```

In [155…
```
dataset.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 [154…
```
dataset['deck'].isnull().value_counts()
```

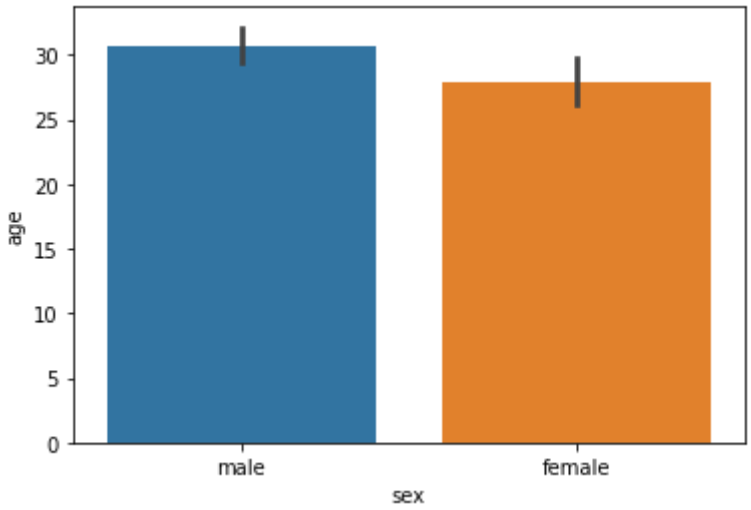
Out[154…
```
True     688
False    203
Name: deck, dtype: int64
```

In [153…
```
dataset['adult_male'].value_counts()
```

Out[153…
```
True     537
False    354
Name: adult_male, dtype: int64
```

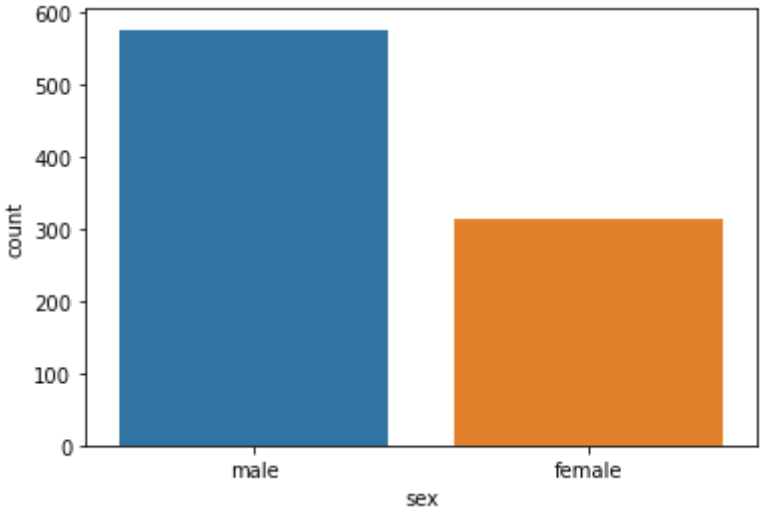
In [145…
```
sns.barplot(x='sex', y='age', data=dataset)
```

Out[145…
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



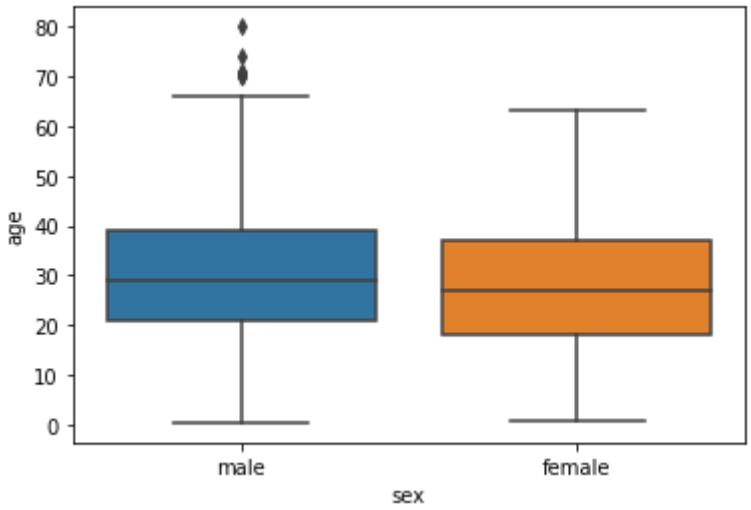
In [146…
```
sns.countplot(x='sex', data=dataset)
```

Out[146…
```
<AxesSubplot:xlabel='sex', ylabel='count'>
```



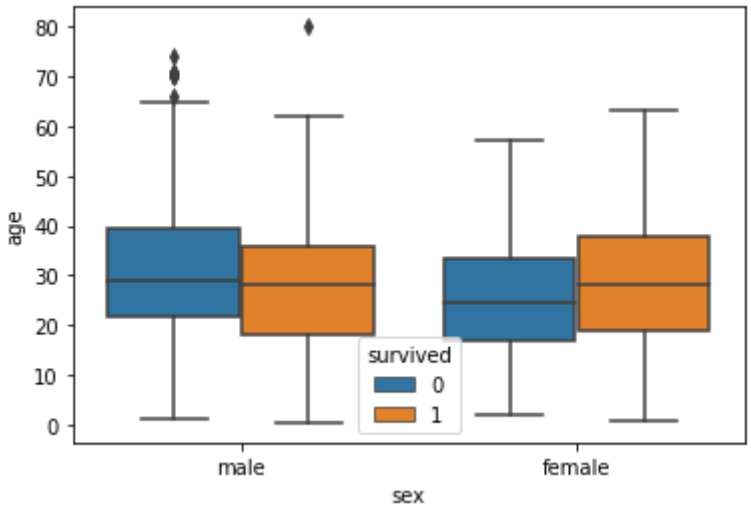
In [147…
```
sns.boxplot(x='sex', y='age', data=dataset)
```

Out[147…
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



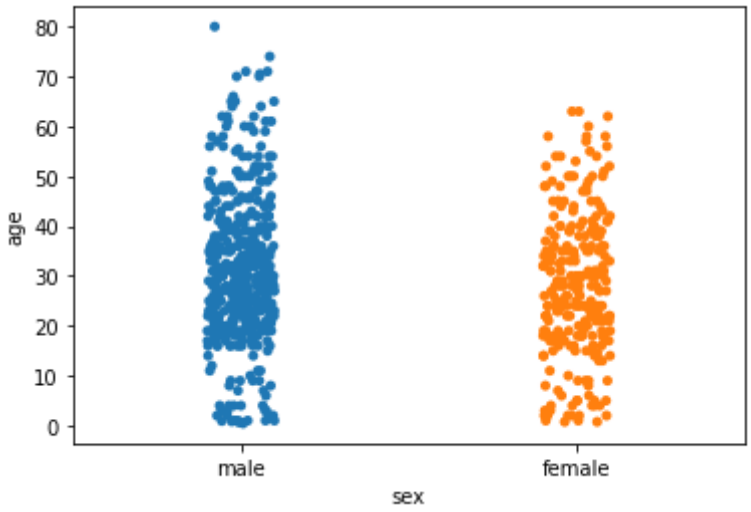
In [148…
```
sns.boxplot(x='sex', y='age', data=dataset, hue="survived")
```

Out[148…
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



In [149…
```
sns.stripplot(x='sex', y='age', data=dataset)
```

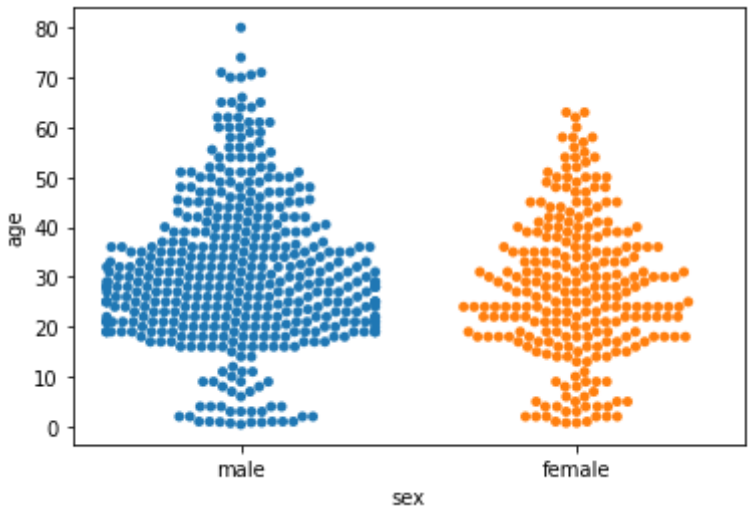
Out[149… <AxesSubplot:xlabel='sex', ylabel='age'>



In [150…
```
sns.swarmplot(x='sex', y='age', data=dataset)
```

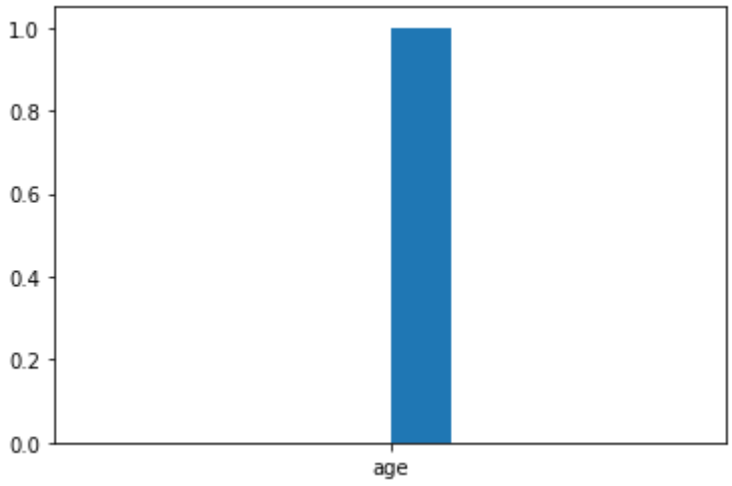
```
/home/pktc-320/anaconda3/lib/python3.9/site-packages/seaborn/categorical.py:1296: UserWarning: 5.9% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
```

Out[150… <AxesSubplot:xlabel='sex', ylabel='age'>



In [151…
```
plt.hist(x='age')
```

Out[151…
```
(array([0., 0., 0., 0., 0., 1., 0., 0., 0., 0.]),
 array([-0.5, -0.4, -0.3, -0.2, -0.1,  0. ,  0.1,  0.2,  0.3,  0.4,  0.5]),
 <BarContainer object of 10 artists>)
```



In [ ]: