

pratical-8

April 17, 2024

1 Data Visualization I

1. 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.
2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

```
[ ]: # # Import required libraries
# import pandas as pd
# import numpy as np
# import matplotlib.pyplot as plt
# import seaborn as sns

# # Load the dataset
# dataset = sns.load_dataset('titanic')
# dataset.head()

# # A. Distribution Plots

# # a. Dist-Plot
# sns.distplot(x = dataset['age'], bins = 10)

# sns.distplot(dataset['age'], bins = 10,kde=False)

# # b. Joint Plot
# # For Plot 1
# sns.jointplot(x = dataset['age'], y = dataset['fare'], kind = 'scatter')

# # For Plot 2
# sns.jointplot(x = dataset['age'], y = dataset['fare'], kind = 'hex')

# # d. Rug Plot
# sns.rugplot(dataset['fare'])
```

```

# # B. Categorical Plots

# # a. Bar Plot
# sns.barplot(x='sex', y='age', data=dataset)

# sns.barplot(x='sex', y='age', data=dataset, estimator=np.std)

# # b. Count Plot
# sns.countplot(x='sex', data=dataset)

# # c. Box Plot
# sns.boxplot(x='sex', y='age', data=dataset)

# sns.boxplot(x='sex', y='age', data=dataset, hue="survived")

# # d. Violin Plot
# sns.violinplot(x='sex', y='age', data=dataset)

# sns.violinplot(x='sex', y='age', data=dataset, hue='survived')

# # C. Advanced Plots

# # a. Strip Plot
# sns.stripplot(x='sex', y='age', data=dataset, jitter=False)

# sns.stripplot(x='sex', y='age', data=dataset, jitter=True, hue='survived')

# # b. Swarm Plot
# sns.swarmplot(x='sex', y='age', data=dataset)

# sns.swarmplot(x='sex', y='age', data=dataset, hue='survived')

# # D. Matrix Plots

# # a. Heat Map
# dataset = dataset[['survived', 'pclass', 'age', 'sibsp', 'parch', 'fare',
# ↪ 'adult_male', 'alone']]
# dataset.corr()
# corr = dataset.corr()
# sns.heatmap(corr)

# sns.heatmap(corr, annot=True)

# # b. Cluster Map
# sns.histplot(dataset['fare'], kde=False, bins=10)

```

```
[1]: # Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[2]: # Load the dataset
dataset = sns.load_dataset('titanic')
dataset.head()
```

```
[2]:   survived  pclass    sex  age  sibsp  parch   fare embarked  class \
0         0        3  male  22.0     1     0   7.2500          S  Third
1         1        1 female  38.0     1     0  71.2833          C  First
2         1        3 female  26.0     0     0   7.9250          S  Third
3         1        1 female  35.0     1     0  53.1000          S  First
4         0        3  male  35.0     0     0   8.0500          S  Third

      who  adult_male deck  embark_town  alive  alone
0    man         True  NaN  Southampton    no  False
1  woman        False    C   Cherbourg   yes  False
2  woman        False  NaN  Southampton   yes   True
3  woman        False    C   Southampton   yes  False
4    man         True  NaN  Southampton    no   True
```

```
[3]: # A. Distribution Plots
# a. Dist-Plot
sns.distplot(x = dataset['age'], bins = 10)
```

C:\Users\Dell\AppData\Local\Temp\ipykernel_7688\724736093.py:3: UserWarning:

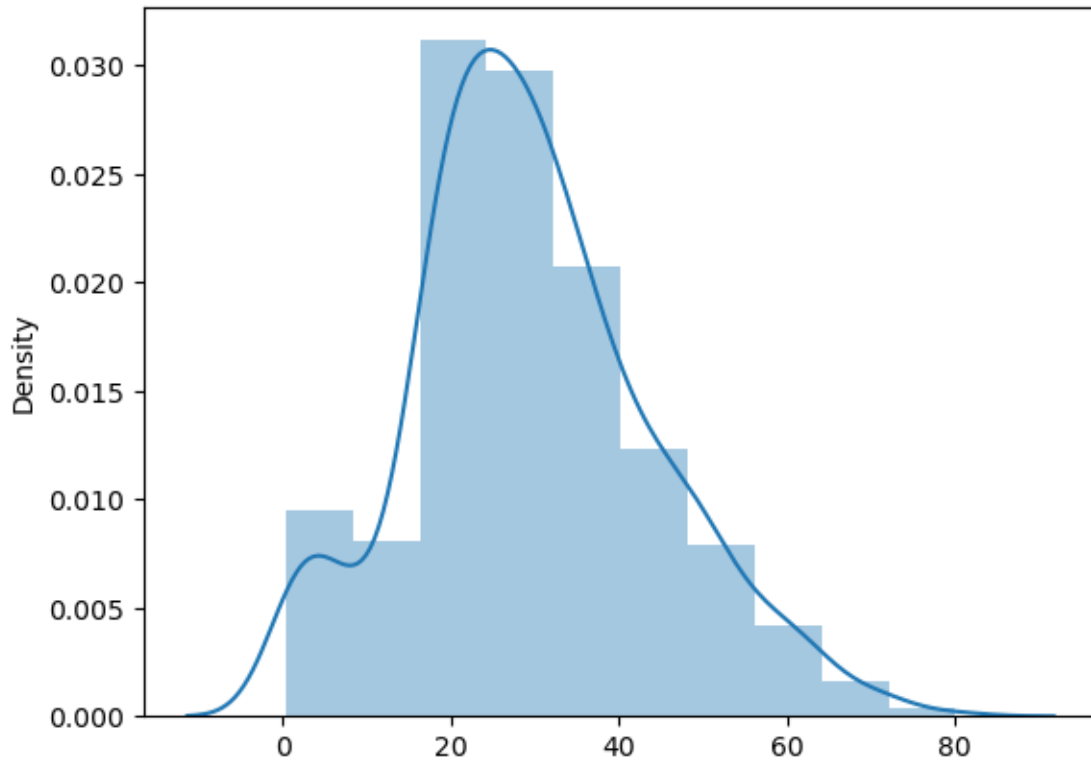
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

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(x = dataset['age'], bins = 10)
F:\Anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

```
[3]: <Axes: ylabel='Density'>
```



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

C:\Users\Dell\AppData\Local\Temp\ipykernel_7688\3517108427.py:1: UserWarning:

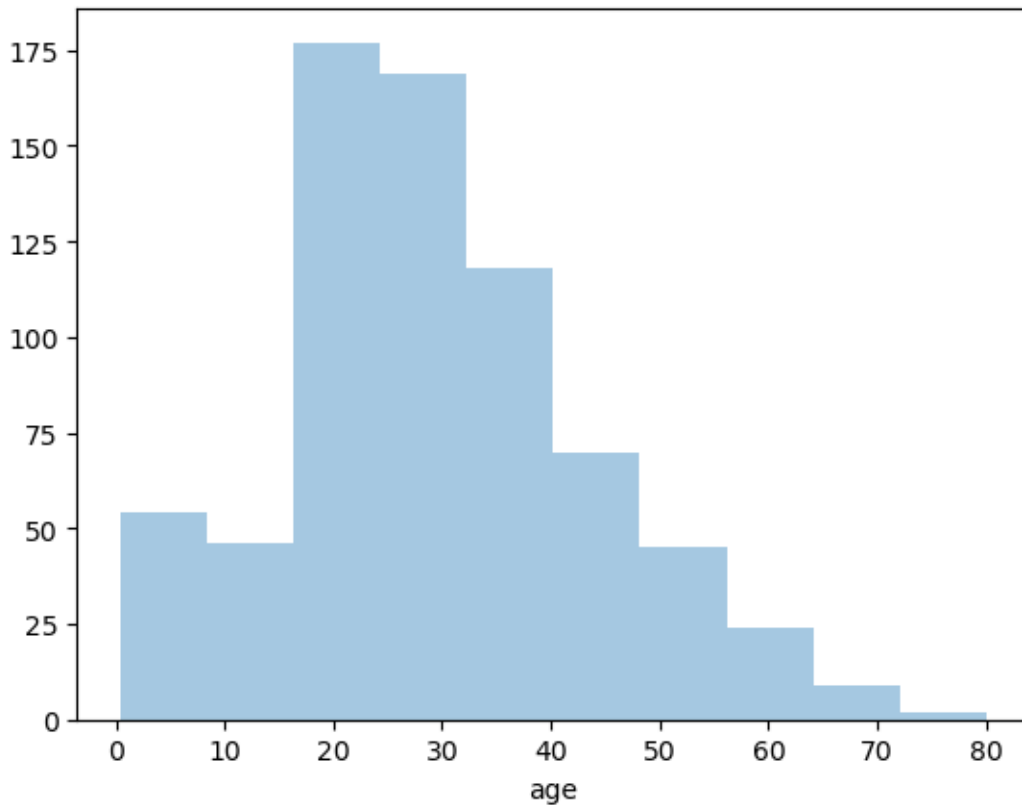
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

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(dataset['age'], bins = 10,kde=False)
```

```
[4]: <Axes: xlabel='age'>
```



```
[5]: # b. Joint Plot
      # For Plot 1
      sns.jointplot(x = dataset['age'], y = dataset['fare'], kind = 'scatter')
```

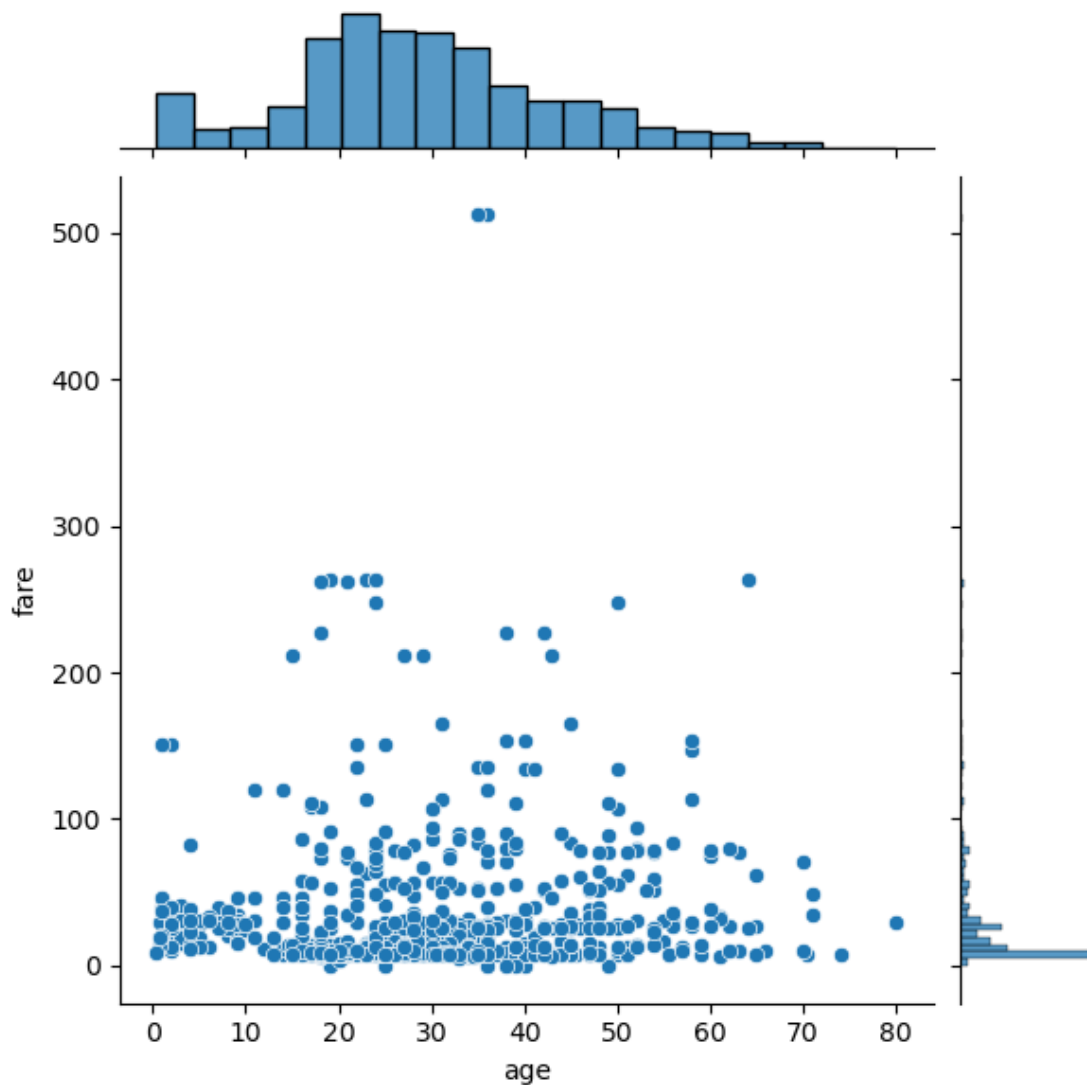
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
[5]: <seaborn.axisgrid.JointGrid at 0x1f191d96c50>
```



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

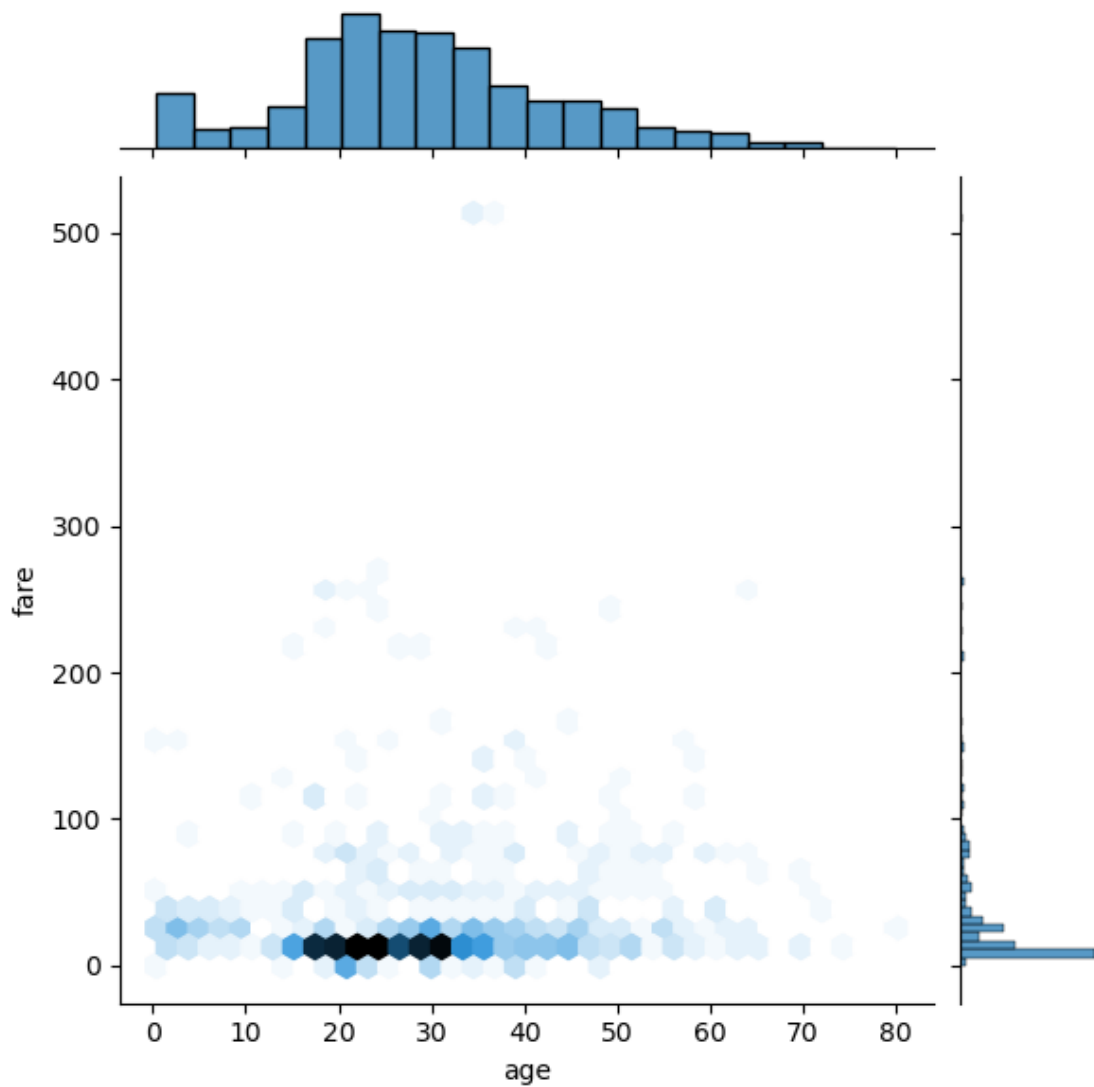
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

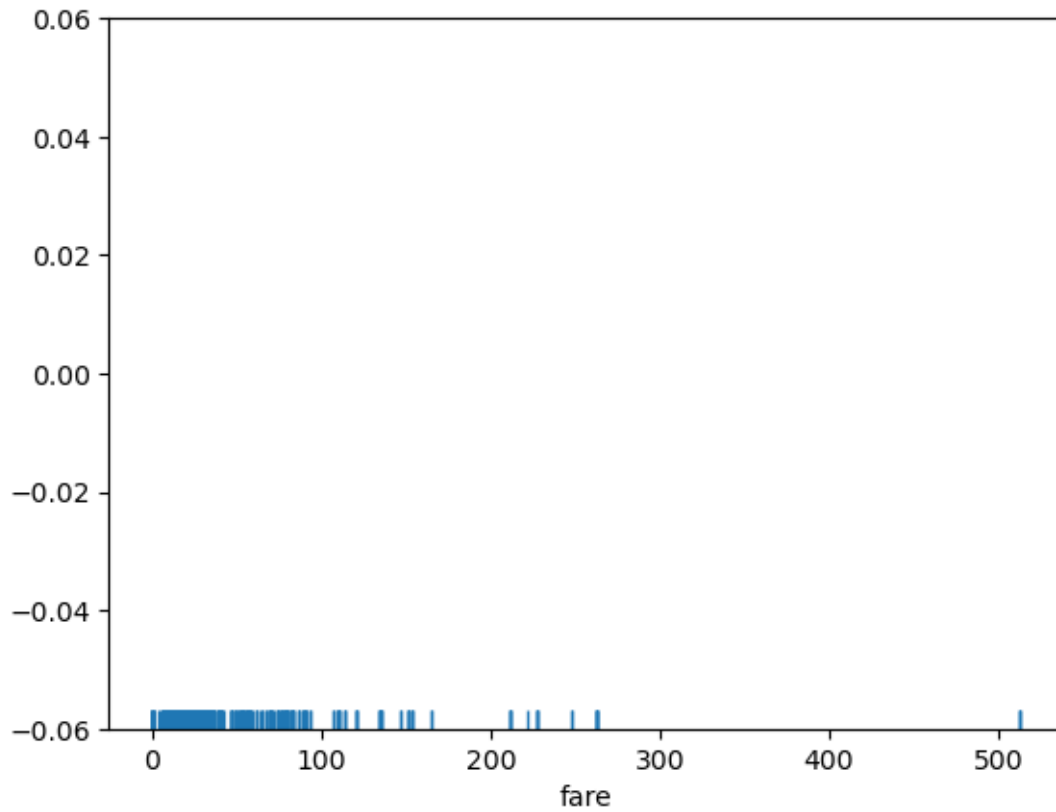
```
[6]: <seaborn.axisgrid.JointGrid at 0x1f191daf6d0>
```



```
[7]: # d. Rug Plot
sns.rugplot(dataset['fare'])
```

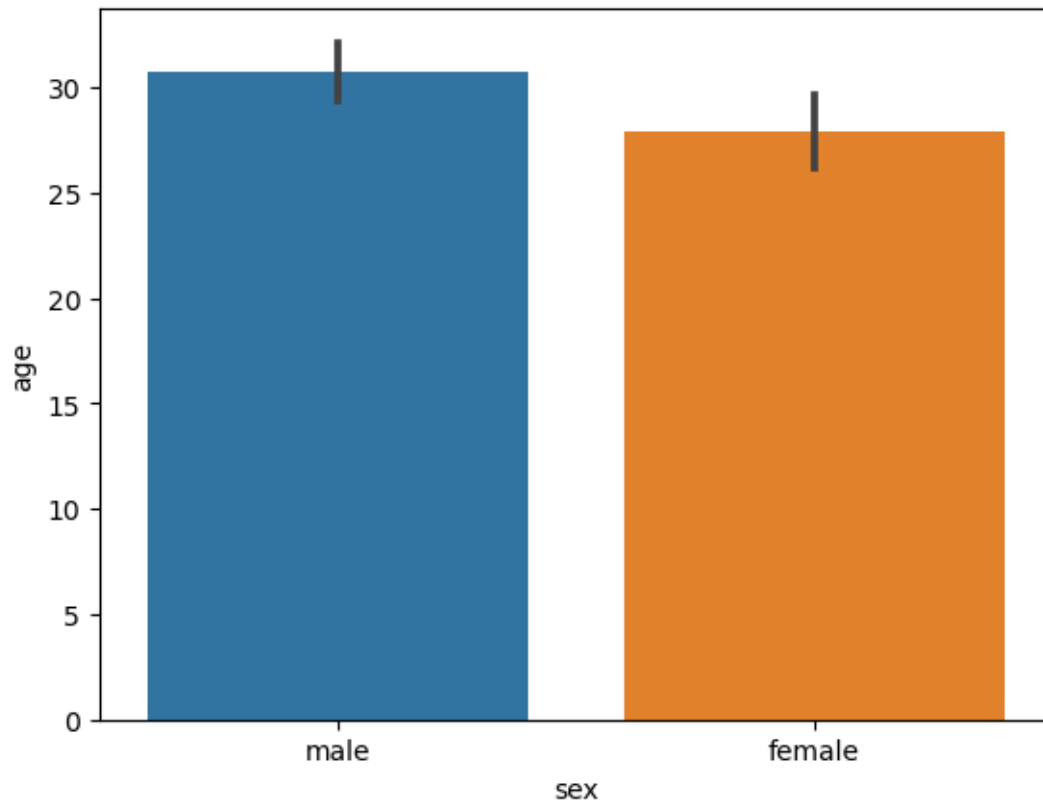
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
with pd.option_context('mode.use_inf_as_na', True):

```
[7]: <Axes: xlabel='fare'>
```



```
[8]: # B. Categorical Plots  
  
# a. Bar Plot  
sns.barplot(x='sex', y='age', data=dataset)
```

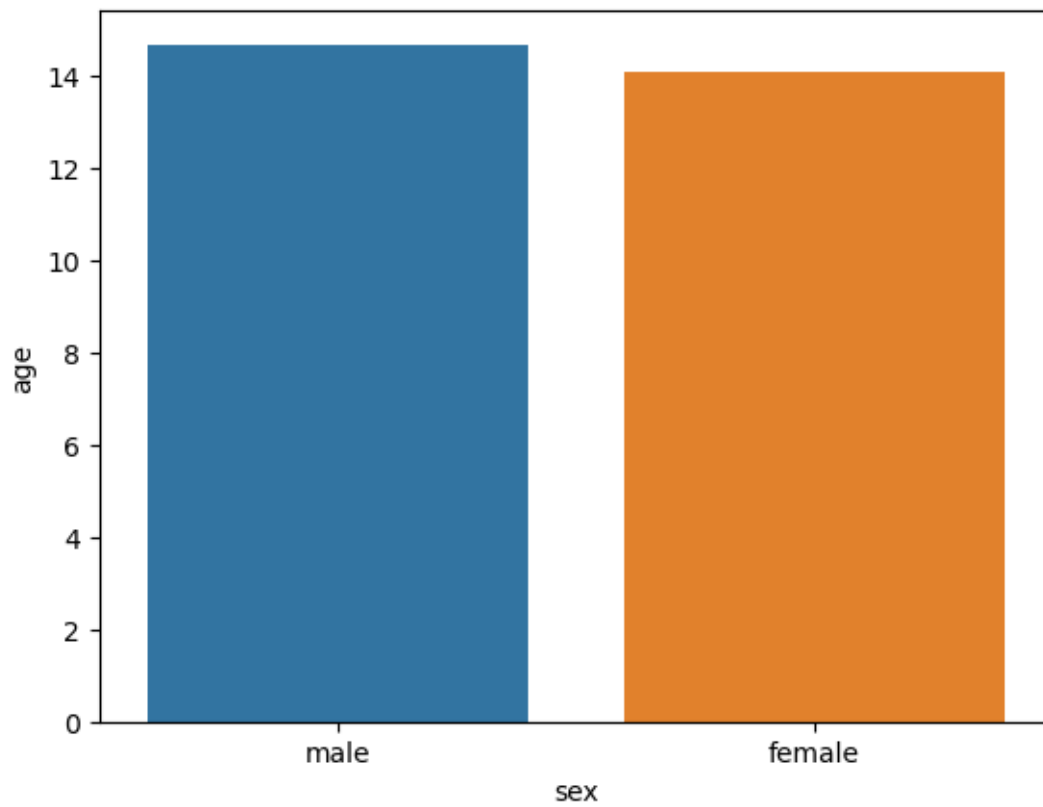
```
[8]: <Axes: xlabel='sex', ylabel='age'>
```

```
[9]: sns.barplot(x='sex', y='age', data=dataset, estimator=np.std)
```

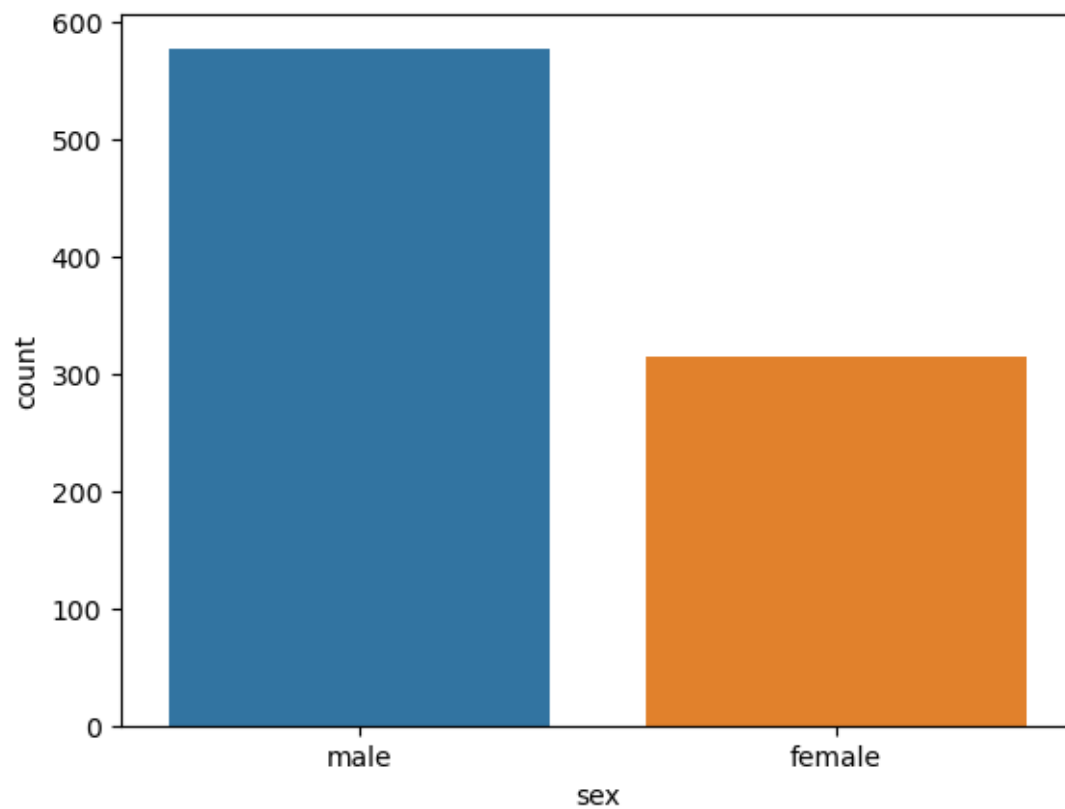
```
F:\Anaconda3\Lib\site-packages\numpy\lib\nanfunctions.py:1556: RuntimeWarning:  
All-NaN slice encountered  
  return function_base._ureduce(a,
```

```
[9]: <Axes: xlabel='sex', ylabel='age'>
```



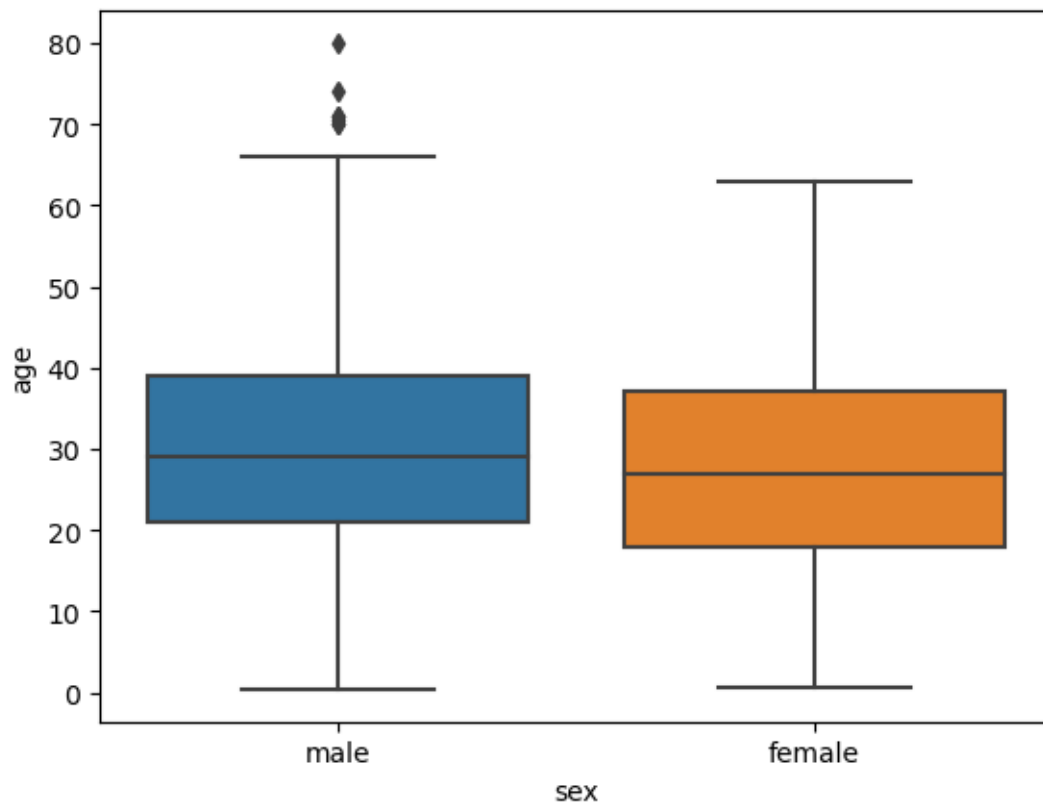
```
[10]: # b. Count Plot
sns.countplot(x='sex', data=dataset)
```

```
[10]: <Axes: xlabel='sex', ylabel='count'>
```



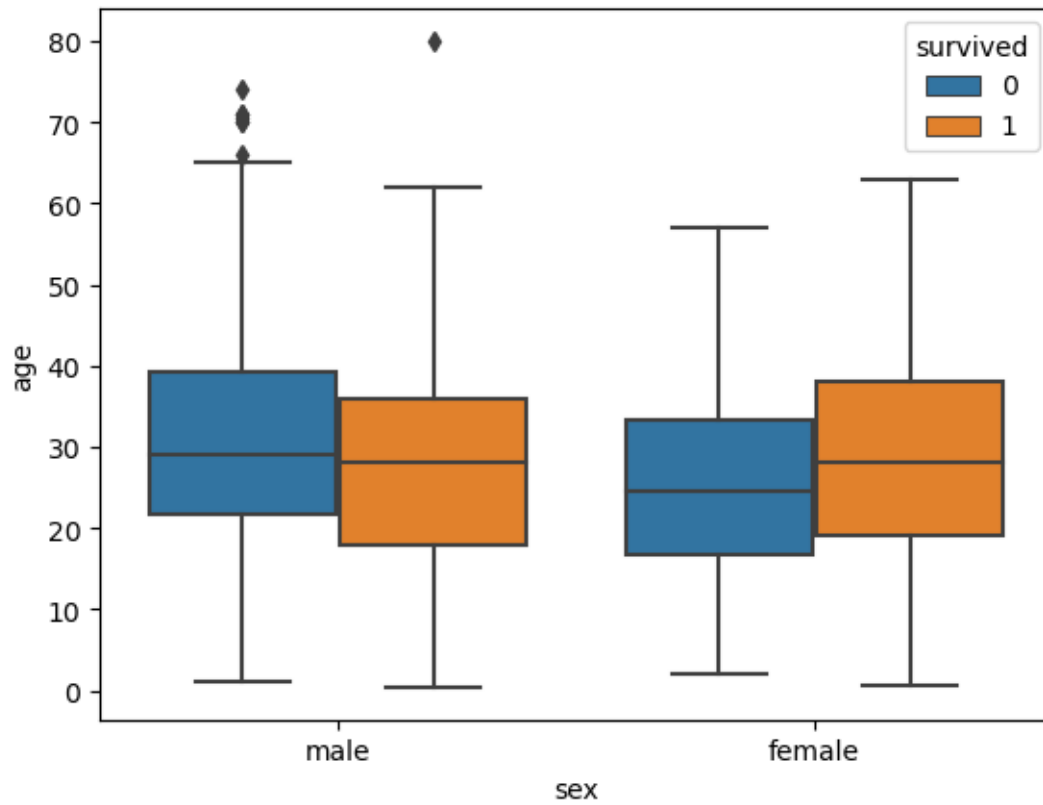
```
[11]: # c. Box Plot  
sns.boxplot(x='sex', y='age', data=dataset)
```

```
[11]: <Axes: xlabel='sex', ylabel='age'>
```



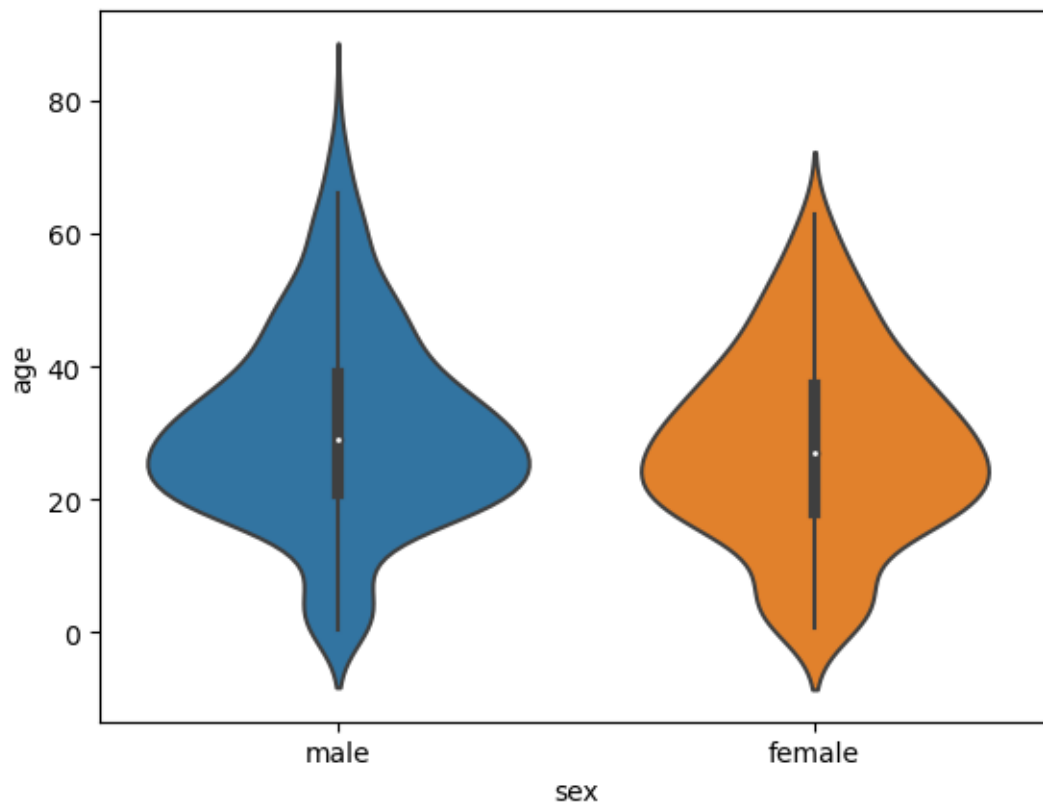
```
[12]: sns.boxplot(x='sex', y='age', data=dataset, hue="survived")
```

```
[12]: <Axes: xlabel='sex', ylabel='age'>
```



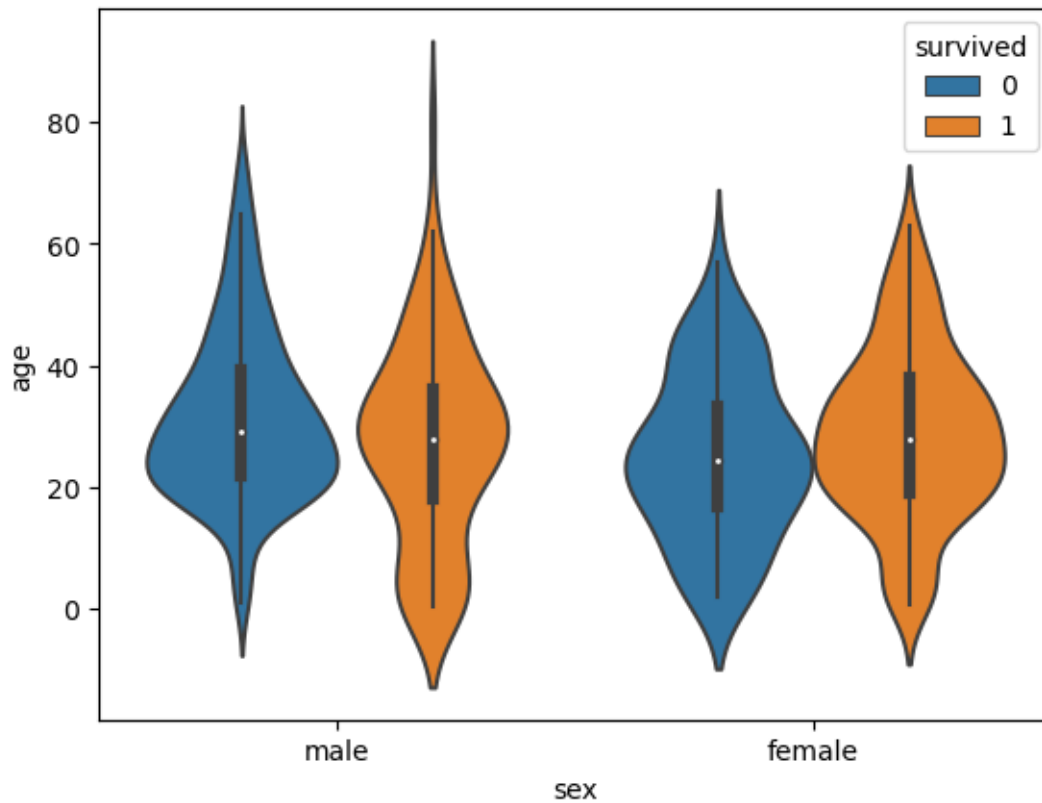
```
[13]: # d. Violin Plot
sns.violinplot(x='sex', y='age', data=dataset)
```

```
[13]: <Axes: xlabel='sex', ylabel='age'>
```



```
[14]: sns.violinplot(x='sex', y='age', data=dataset, hue='survived')
```

```
[14]: <Axes: xlabel='sex', ylabel='age'>
```



```
[15]: # C. Advanced Plots
```

```
# a. Strip Plot
```

```
sns.stripplot(x='sex', y='age', data=dataset, jitter=False)
```

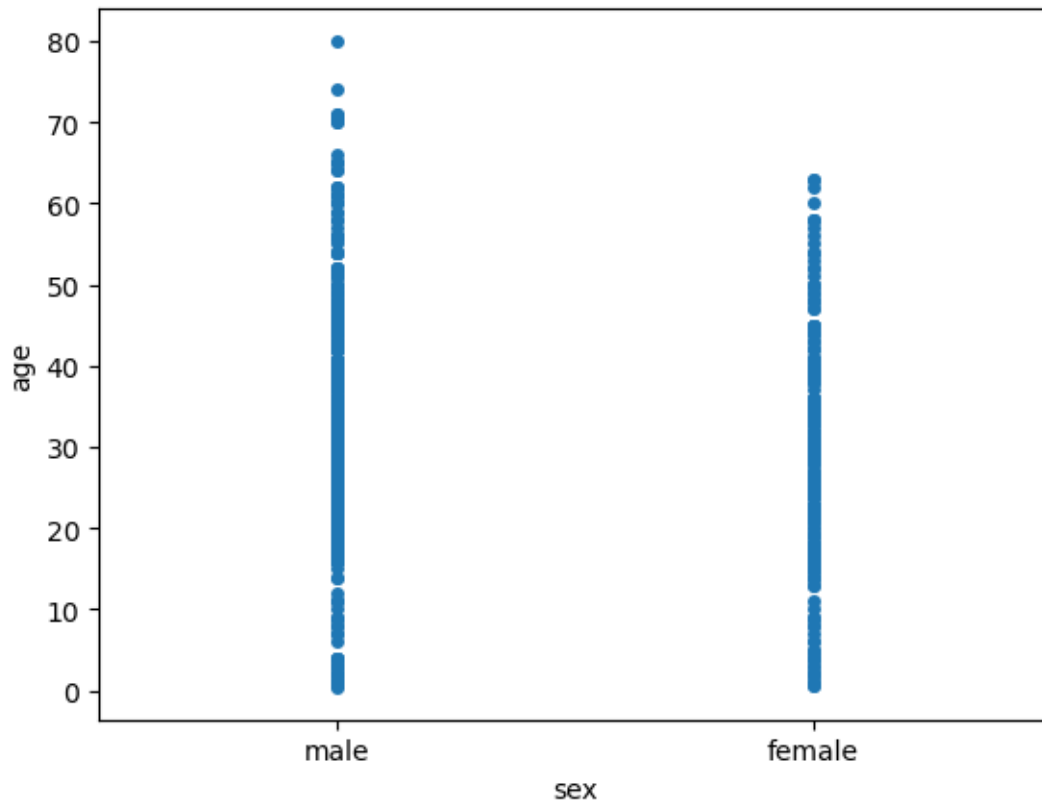
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
[15]: <Axes: xlabel='sex', ylabel='age'>
```



```
[16]: sns.stripplot(x='sex', y='age', data=dataset, jitter=True, hue='survived')
```

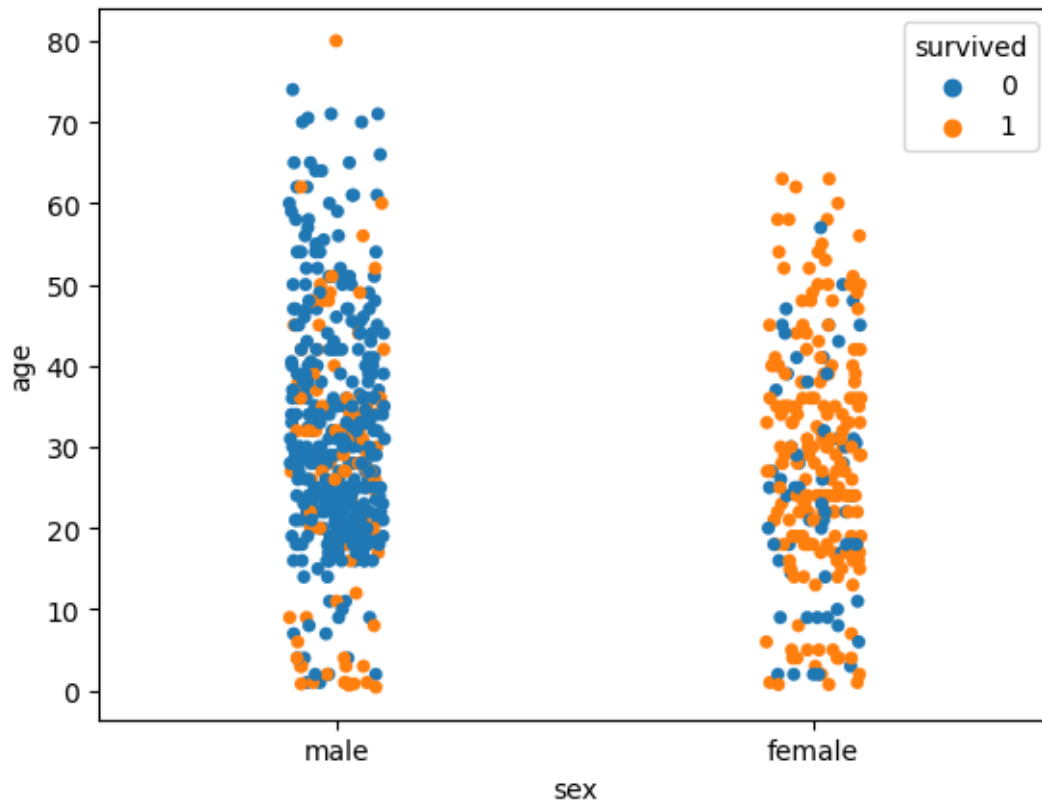
```
F:\Anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
```

```
    with pd.option_context('mode.use_inf_as_na', True):
```

```
F:\Anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
```

```
    with pd.option_context('mode.use_inf_as_na', True):
```

```
[16]: <Axes: xlabel='sex', ylabel='age'>
```

```
[17]: # b. Swarm Plot
sns.swarmplot(x='sex', y='age', data=dataset)
```

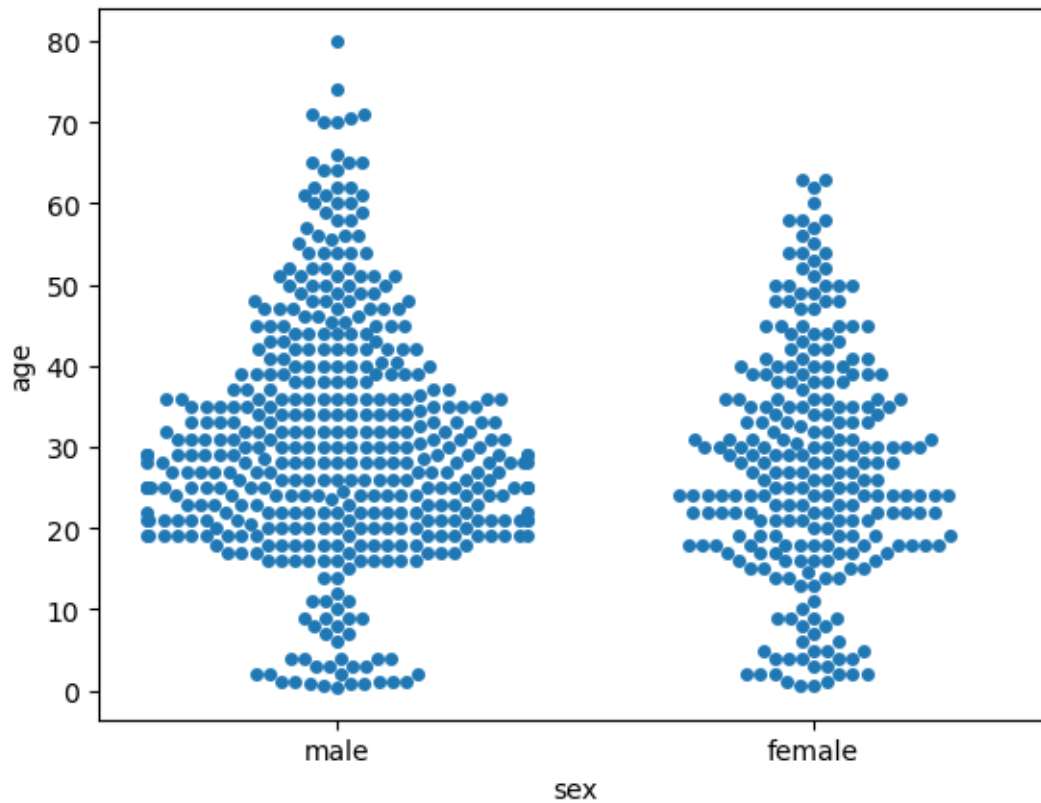
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
[17]: <Axes: xlabel='sex', ylabel='age'>
```



```
[18]: sns.swarmplot(x='sex', y='age', data=dataset, hue='survived')
```

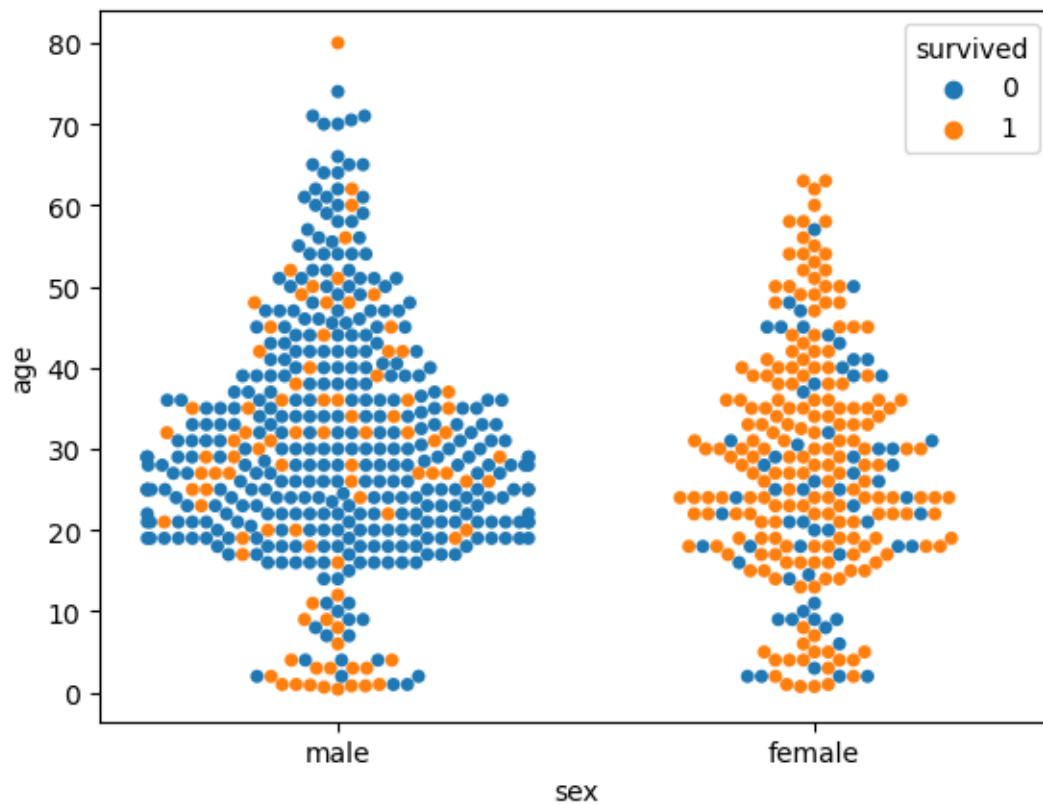
F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

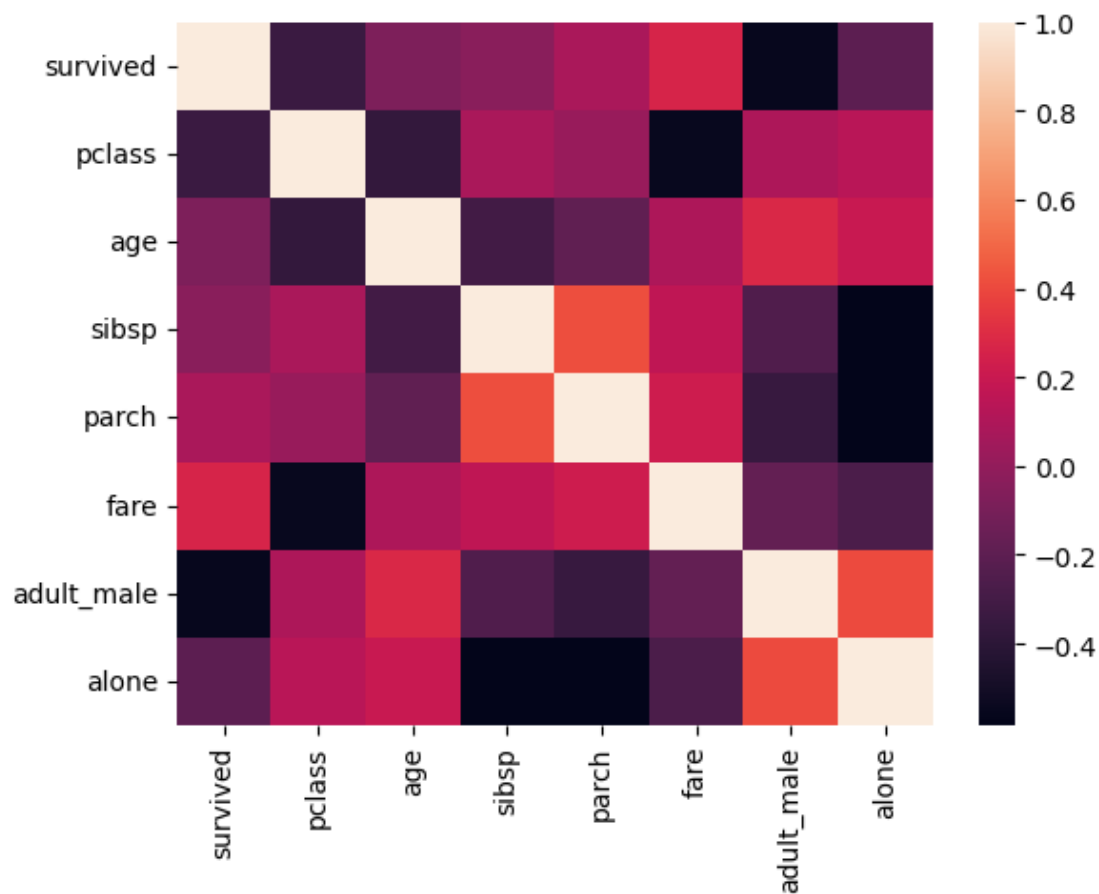
```
with pd.option_context('mode.use_inf_as_na', True):
```

```
[18]: <Axes: xlabel='sex', ylabel='age'>
```



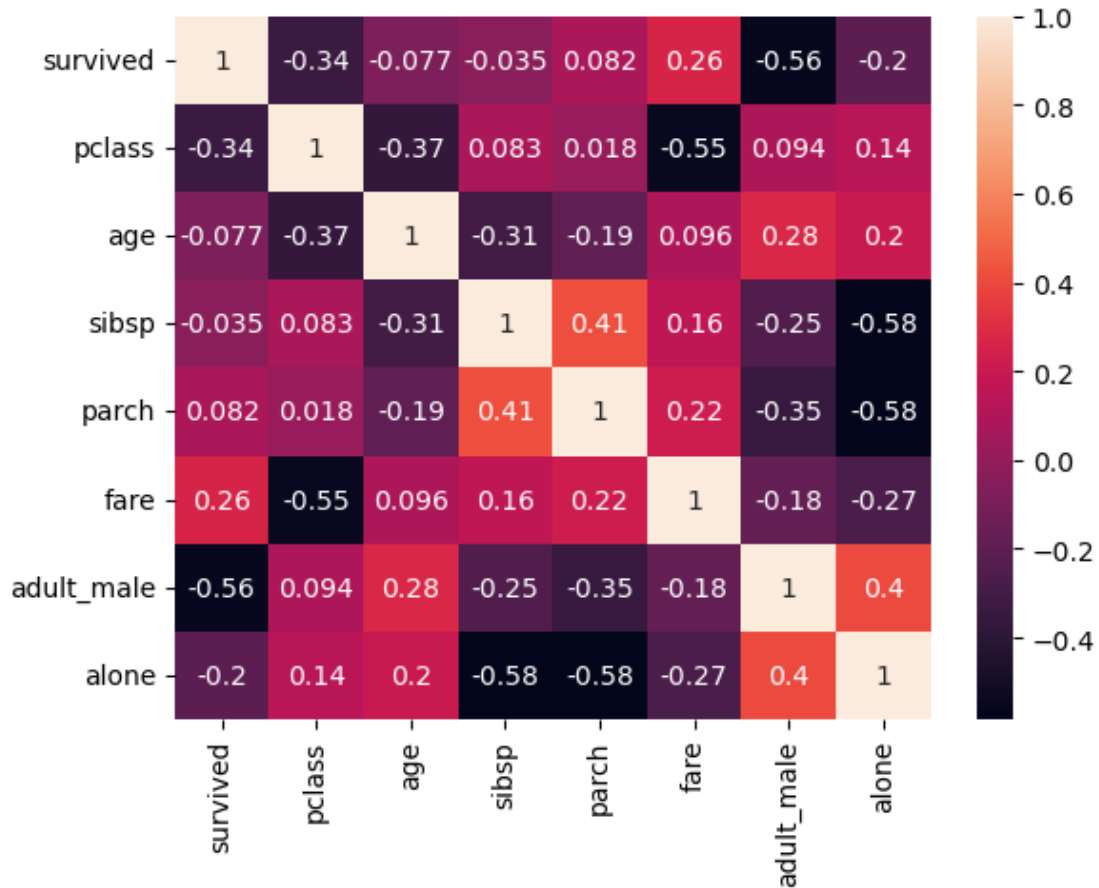
```
[19]: dataset = dataset[['survived', 'pclass', 'age', 'sibsp', 'parch', 'fare',  
    ↪ 'adult_male', 'alone']]  
corr = dataset.corr()  
sns.heatmap(corr)
```

```
[19]: <Axes: >
```



```
[20]: sns.heatmap(corr, annot=True)
```

```
[20]: <Axes: >
```

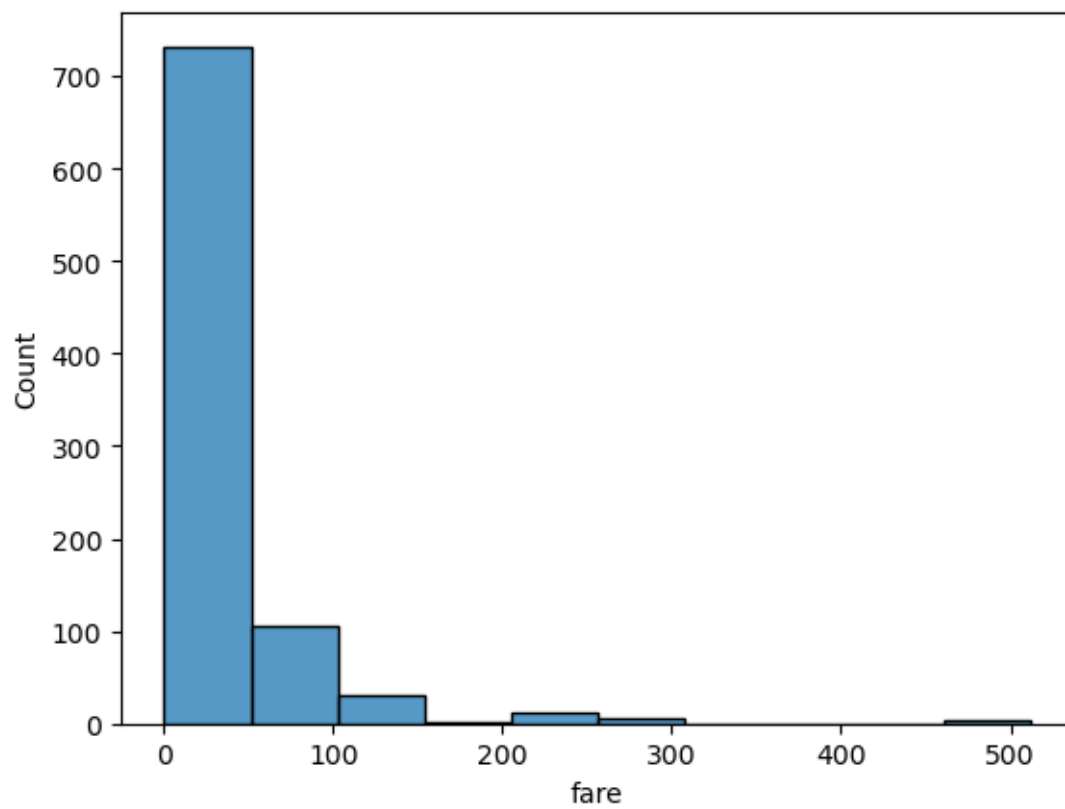


```
[21]: # b. Cluster Map
sns.histplot(dataset['fare'], kde=False, bins=10)
```

F:\Anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

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



[]: