Data Visualization 9

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

In [2]:
```
df = sns.load_dataset('titanic')
```

In [4]:
```
type(df)
```

Out[4]:
```
pandas.core.frame.DataFrame
```

In [5]:
```
df
```

Out[5]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	emb
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Sou
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	C
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Sou
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Sou
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Sou
...	...	...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Sou
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Sou
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Sou
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	C
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Qu

891 rows × 15 columns

In [7]:
```
df.drop('deck',axis=1,inplace=True)
```

In [8]:
```
df
```

Out[8]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	embark_tow
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	Southampt
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	Cherbou
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	Southampt
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	Southampt
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	Southampt
...	...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	Southampt
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	Southampt
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	Southampt
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	Cherbou

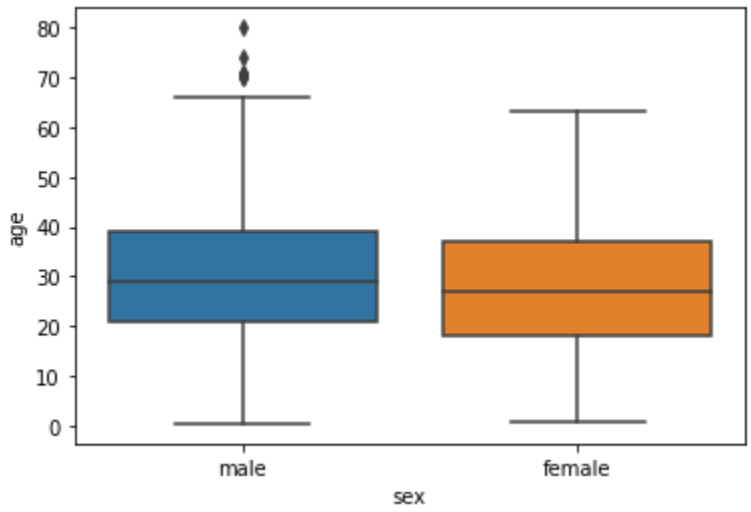
890	0	3	male	32.0	0	0	7.7500		Q	Third	man	True	Queensto

891 rows × 14 columns

BOX PLOT

In [9]:
```
sns.boxplot(x='sex',y='age',data=df)
```

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



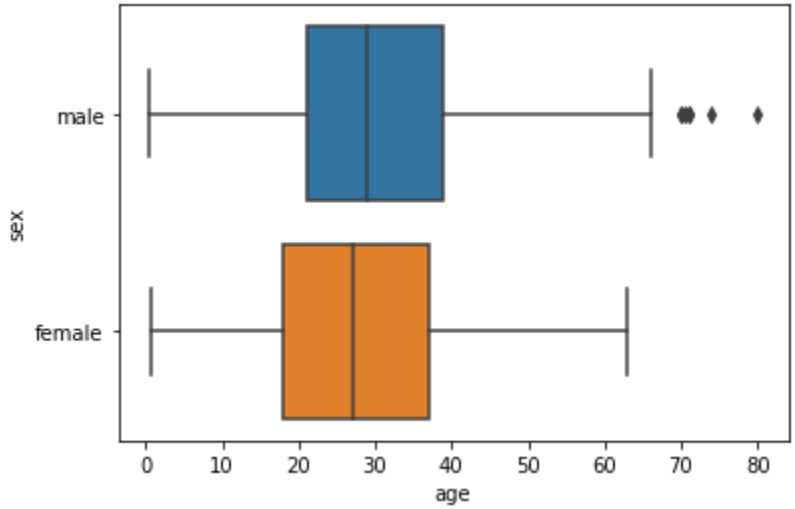
In [10]:
```
## Box plots visually
#show the distribution of numerical data and skewness through displaying the
#data quartiles (or percentiles) and averages.
```

INFERENCE

- Minimum age is 0
- 25th-50th percentile people are of age from 20-30
- 50th-75th percentile people are of age from 30-40
- above 75th percentile people are of age abve 40
- With hue we can also check the age group of people survived and not survived

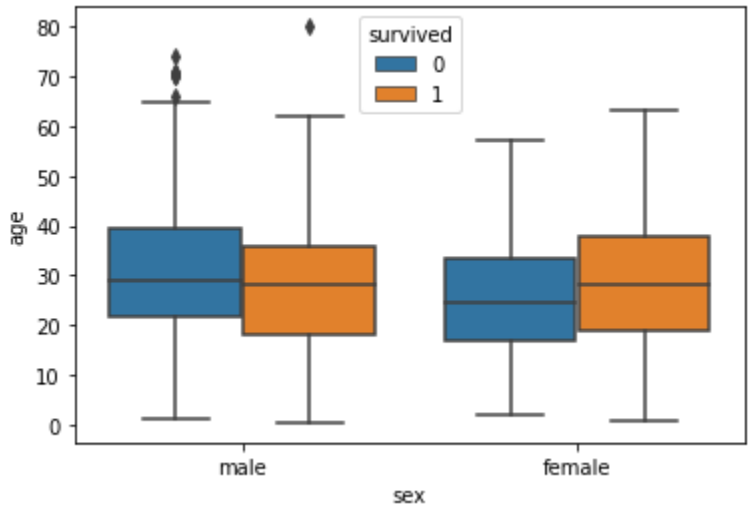
In [11]:
```
sns.boxplot(x='age',y='sex',data=df)
```

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



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

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



Inference: In above chart ,in addition to the information about the age of each gender, you can also see the distribution of the passengers who survived. For instance, you can see that among the male passengers, on average more younger people survived as compared to the older ones.

Inter Quartile Range

In [16]:
```
df_age_sex = df[['age','sex']]
```

In [17]:
```
df_age_sex
```

Out[17]:

	age	sex
0	22.0	male
1	38.0	female
2	26.0	female
3	35.0	female
4	35.0	male
...	...	...
886	27.0	male
887	19.0	female

888	NaN	female
889	26.0	male
890	32.0	male

891 rows × 2 columns

In [31]:
```
Q1_male = df_age_sex[df_age_sex['sex']=='male']['age'].quantile(0.25)
```

In [32]:
```
Q1_female=df_age_sex[df_age_sex['sex']=='female']['age'].quantile(0.25)
```

In [33]:
```
Q3_male = df_age_sex[df_age_sex['sex']=='male']['age'].quantile(0.75)
```

In [34]:
```
Q3_female =df_age_sex[df_age_sex['sex']=='female']['age'].quantile(0.75)
```

In [35]:
```
IQR = Q3_male-Q1_male
IQR
```

Out[35]:
```
18.0
```

In [36]:
```
IQR = Q3_female-Q1_female
IQR
```

Out[36]:
```
19.0
```

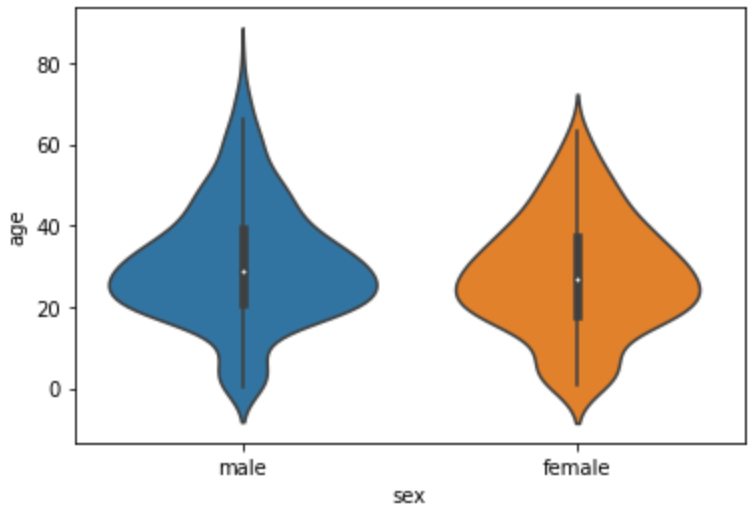
Violin Plot

The 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.

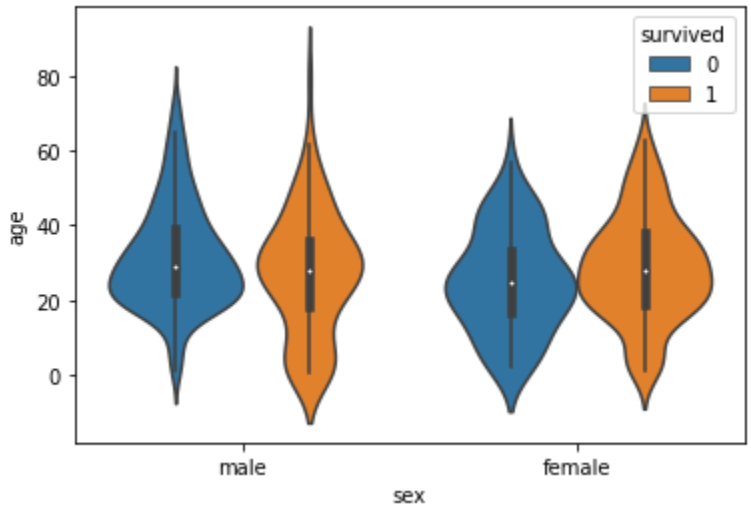
In [37]:
```
sns.violinplot(x='sex',y='age',data=df)
```

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



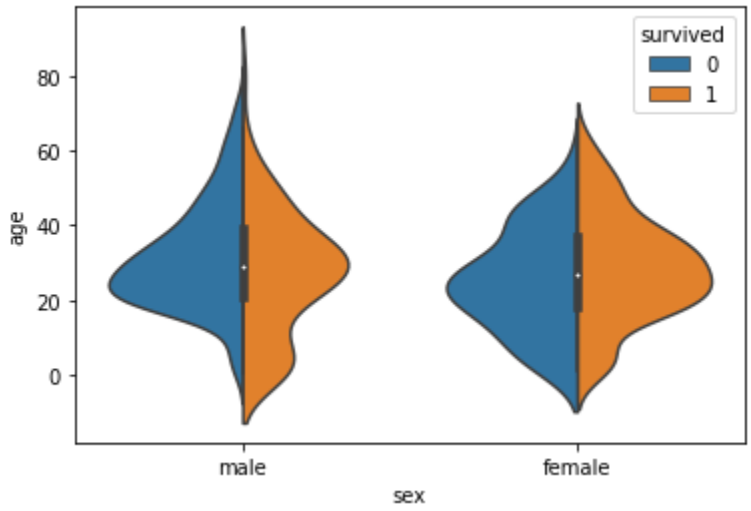
In [38]:
```
sns.violinplot(x='sex',y='age',hue= 'survived',data=df)
```

Out[38]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



In [39]:
```
sns.violinplot(x='sex',y='age',hue= 'survived',split = True,data=df)
```

Out[39]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



Inference: For instance, if you look at the bottom of the violin plot for the males who survived (left-orange), you can see that it is thicker than the bottom of the violin plot for the males who didn't survive (left-blue).

This means that the number of young male passengers who survived is greater than the number of young male passengers who did not survive.

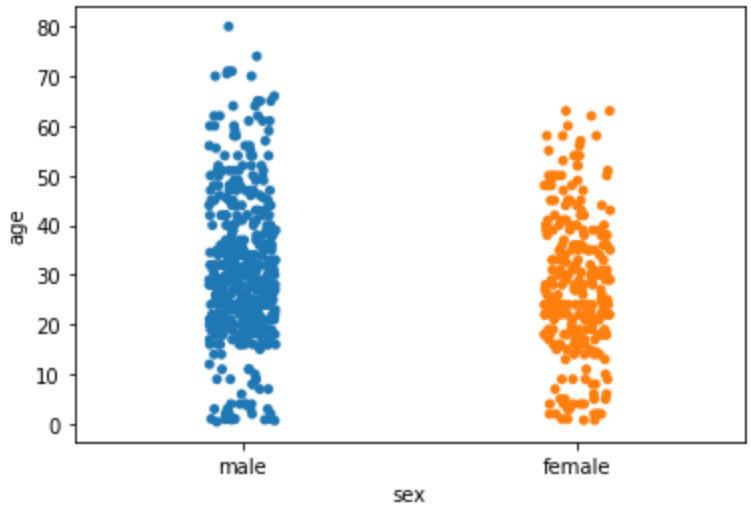
STRIPPLOT

The swarm plot is a combination of the strip and the violin plots.

In the swarmplots, the points are adjusted in such a way that they don't overlap.

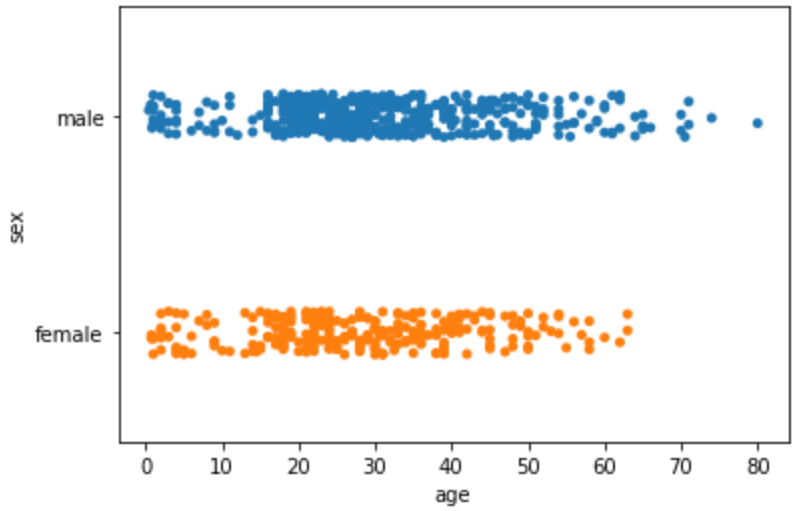
In [52]:
```
sns.stripplot(x='sex',y='age',data=df)
```

Out[52]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



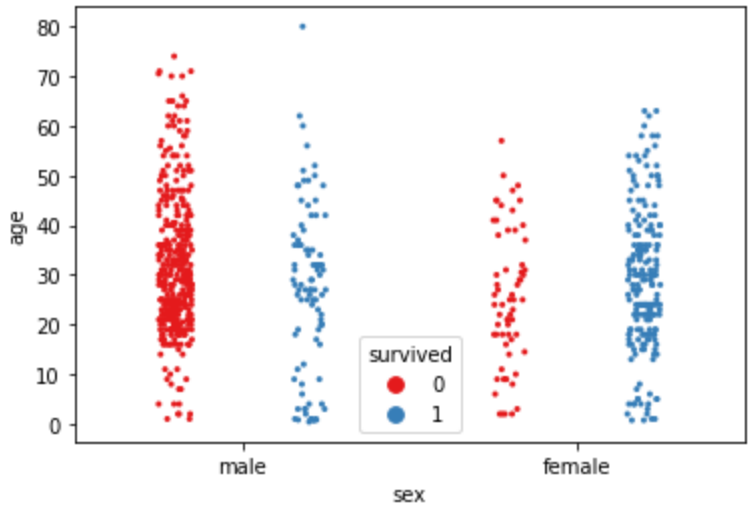
In [53]:
```
sns.stripplot(x='age',y='sex',data=df)
```

Out[53]:
```
<AxesSubplot:xlabel='age', ylabel='sex'>
```



In [54]:
```
sns.stripplot(x='sex',y='age',dodge = True,hue = 'survived',size = 3,palette='Set1',data
```

Out[54]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



Swarm Plot

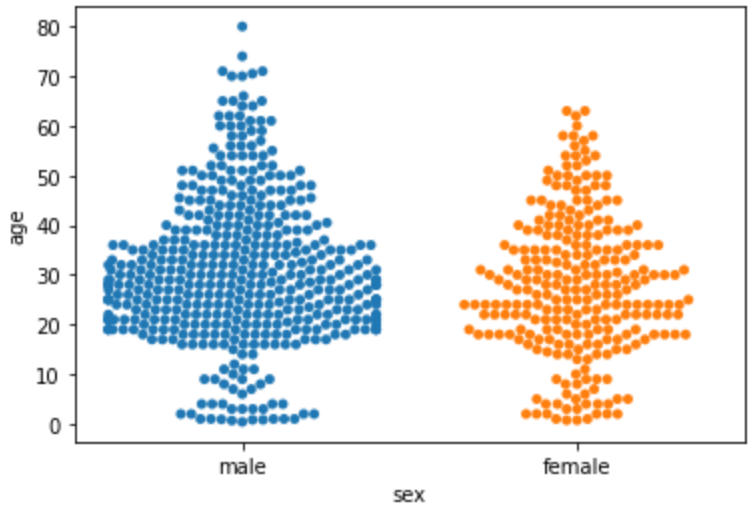
The swarm plot is a combination of the strip and the violin plots.

In the swarmplots, the points are adjusted in such a way that they don't overlap.

In [55]:
```
sns.swarmplot(x='sex',y='age',data=df)
```

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

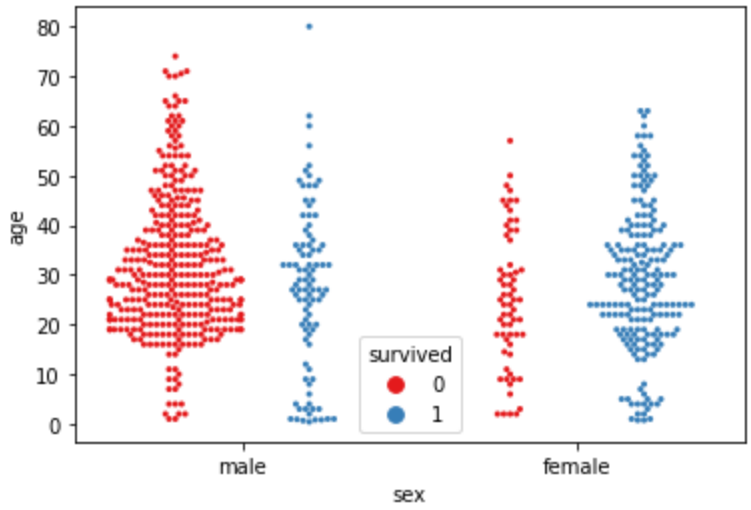
Out[55]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



In [ ]:

In [56]:
```
sns.swarmplot(x='sex',y='age',dodge = True,hue = 'survived',size = 3,palette='Set1',data
```

Out[56]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



if you look at the bottom of the violin plot for the males who survived (red), you can see that it is thicker than the bottom of the violin plot for the males who didn't survive (blue).

In [ ]:

In [ ]:

In [ ]: