

In [1]:

▶

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

▶

```
df = sns.load_dataset('titanic')
```

In [3]:

▶

```
df.head()
```

Out[3]:

	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
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
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

In [4]:

▶

```
df.tail()
```

Out[4]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
886	0	2	male	27.0	0	0	13.00	S	Second	man	
887	1	1	female	19.0	0	0	30.00	S	First	woman	False
888	0	3	female	NaN	1	2	23.45	S	Third	woman	False
889	1	1	male	26.0	0	0	30.00	C	First	man	
890	0	3	male	32.0	0	0	7.75	Q	Third	man	

In [5]:

```
df.describe()
```

Out[5]:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [6]:

```
df.dtypes
```

Out[6]:

```
survived      int64
pclass        int64
sex           object
age           float64
sibsp         int64
parch         int64
fare          float64
embarked      object
class         category
who           object
adult_male    bool
deck         category
embark_town   object
alive         object
alone         bool
dtype: object
```

In [7]:



```
df.isna().sum()
```

Out[7]:

```
survived      0
pclass        0
sex           0
age          177
sibsp         0
parch         0
fare          0
embarked      2
class         0
who           0
adult_male    0
deck         688
embark_town   2
alive         0
alone        0
dtype: int64
```

In [8]:



```
df.shape
```

Out[8]:

```
(891, 15)
```

In [9]:



```
df = df.dropna(subset = 'age')
```

In [10]:



```
df.shape
```

Out[10]:

```
(714, 15)
```

In []:



```
sns.histplot(data = df, y = 'age', x = 'survived')
```

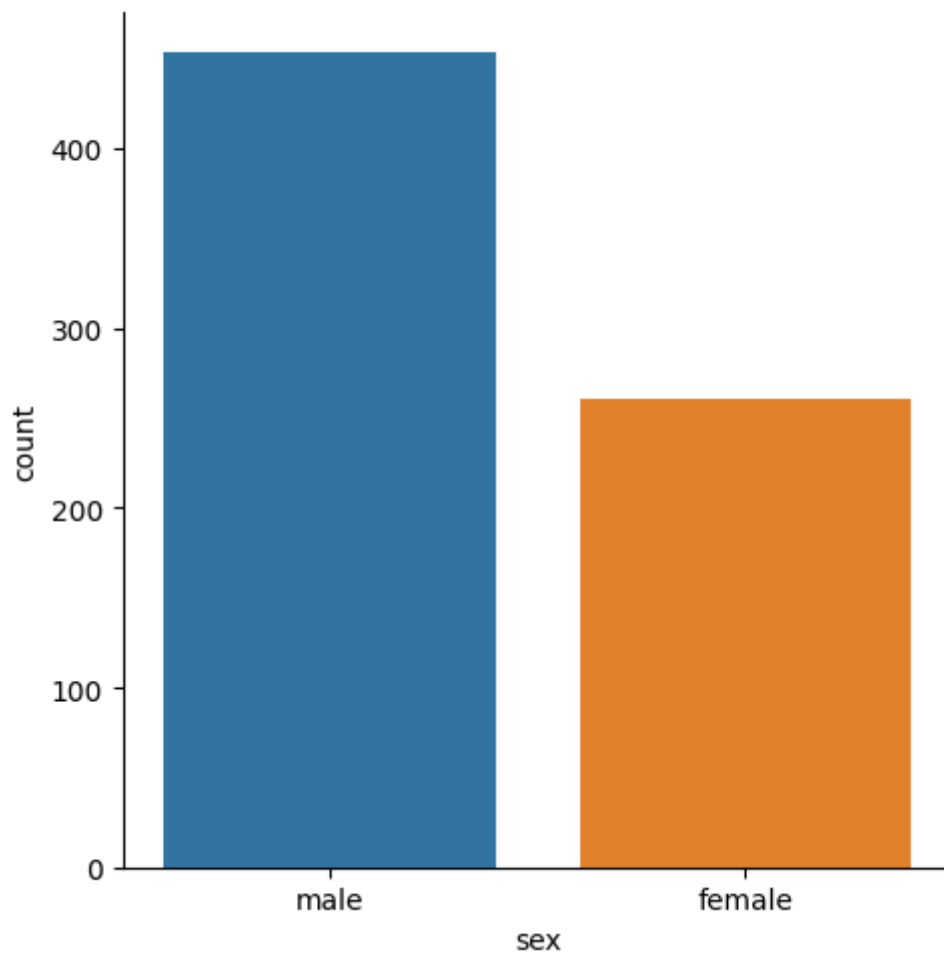
In [58]:



```
sns.catplot(x='sex',data=df,kind='count')
```

Out[58]:

```
<seaborn.axisgrid.FacetGrid at 0x207d7e2eeb0>
```



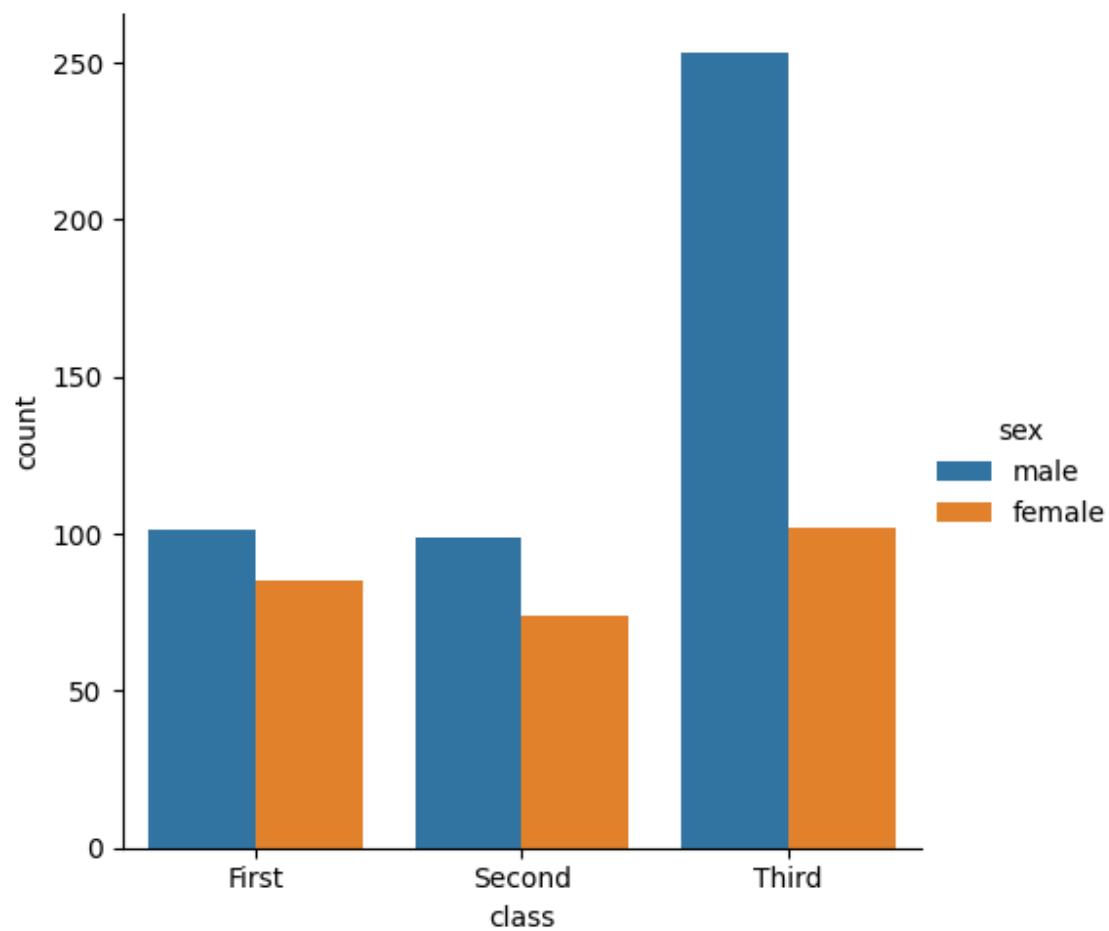
In [47]:



```
sns.factorplot('class',data=df,hue='sex',kind='count')
```

Out[47]:

<seaborn.axisgrid.FacetGrid at 0x207d5d3e340>



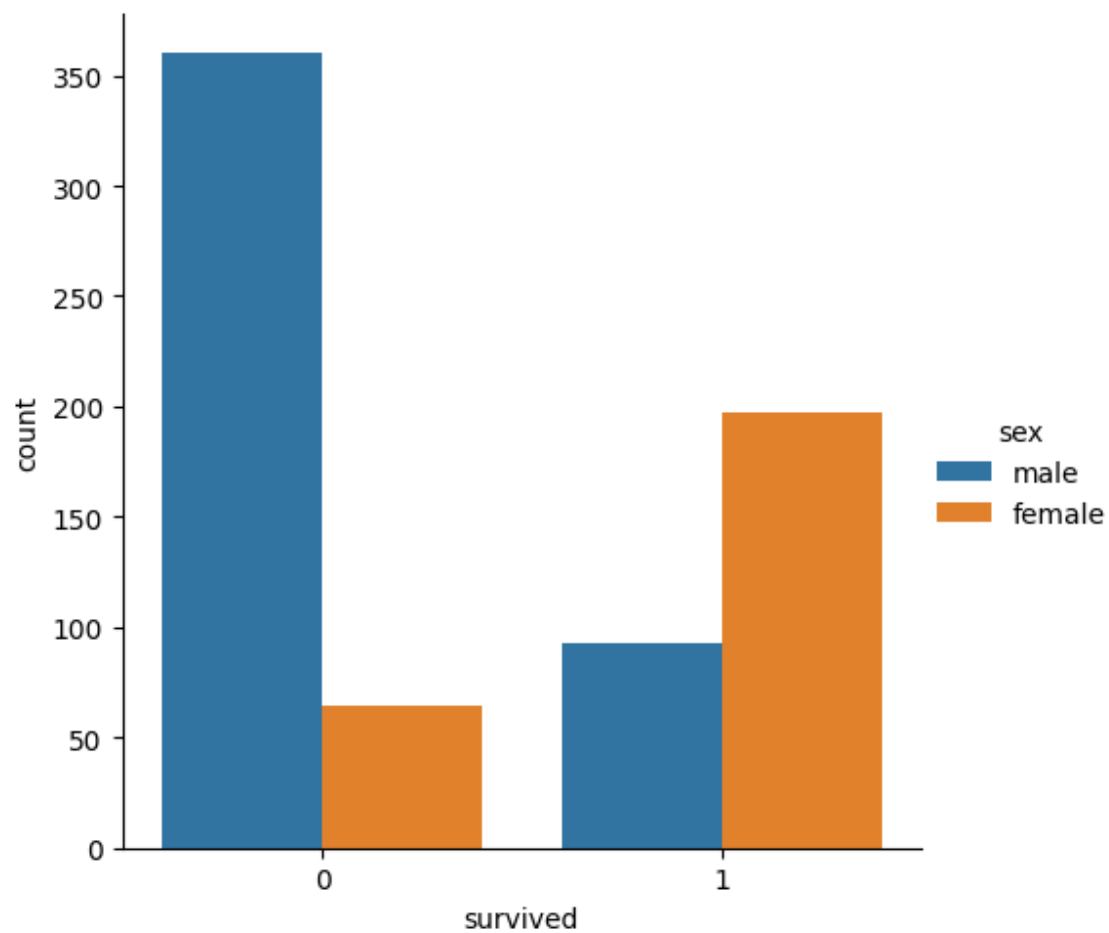
In [51]:



```
sns.catplot(x='survived', data=df, hue='sex', kind='count')
```

Out[51]:

```
<seaborn.axisgrid.FacetGrid at 0x207d688b280>
```

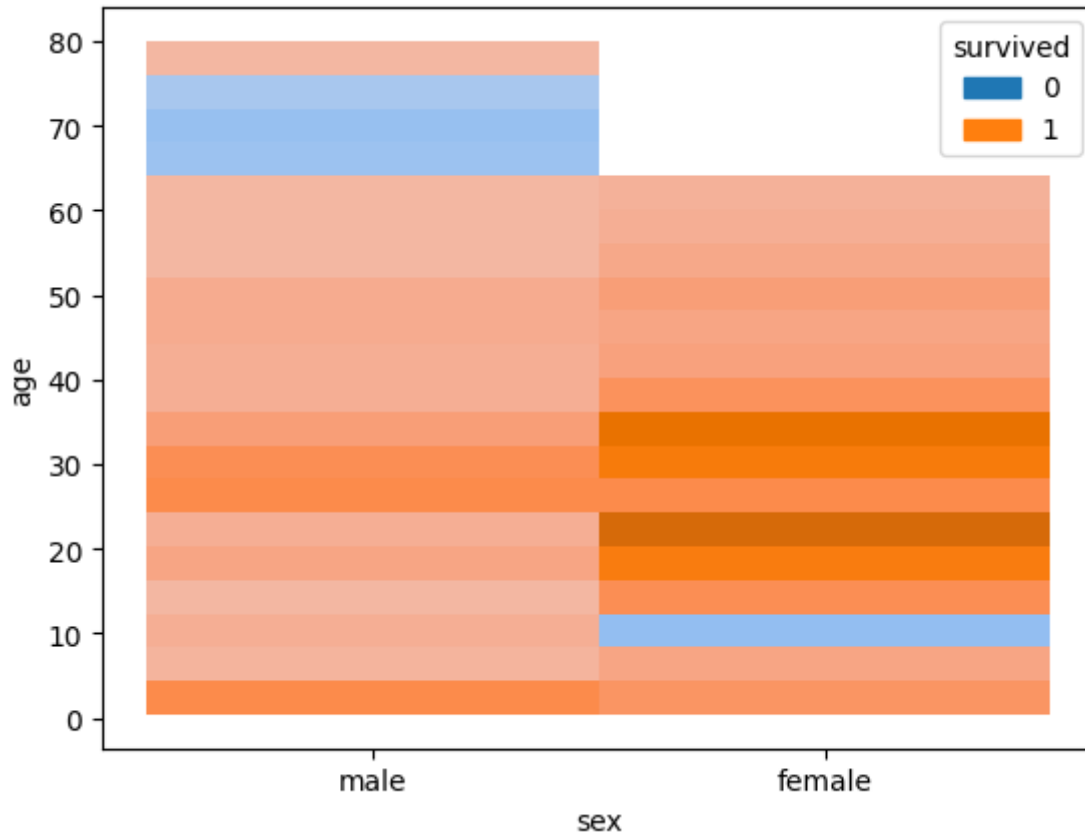


In [49]:

```
sns.histplot(data = df,x='sex',y='age',hue = 'survived')
```

Out[49]:

<Axes: xlabel='sex', ylabel='age'>



In [52]:

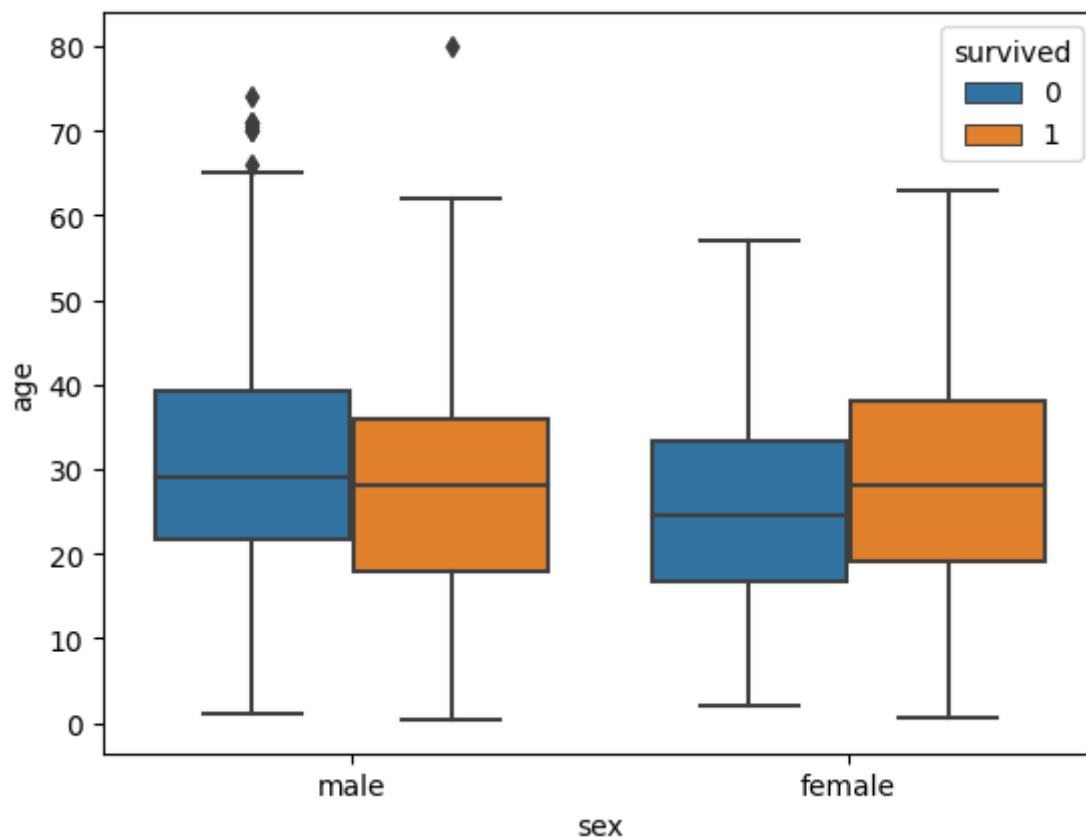
```
df1 = df.copy()  
df1['sex'] = df['sex'].map({'female': 0, 'male': 1})
```

In [53]:

```
sns.boxplot(data = df,x='sex',y='age',hue = 'survived')
```

Out[53]:

<Axes: xlabel='sex', ylabel='age'>



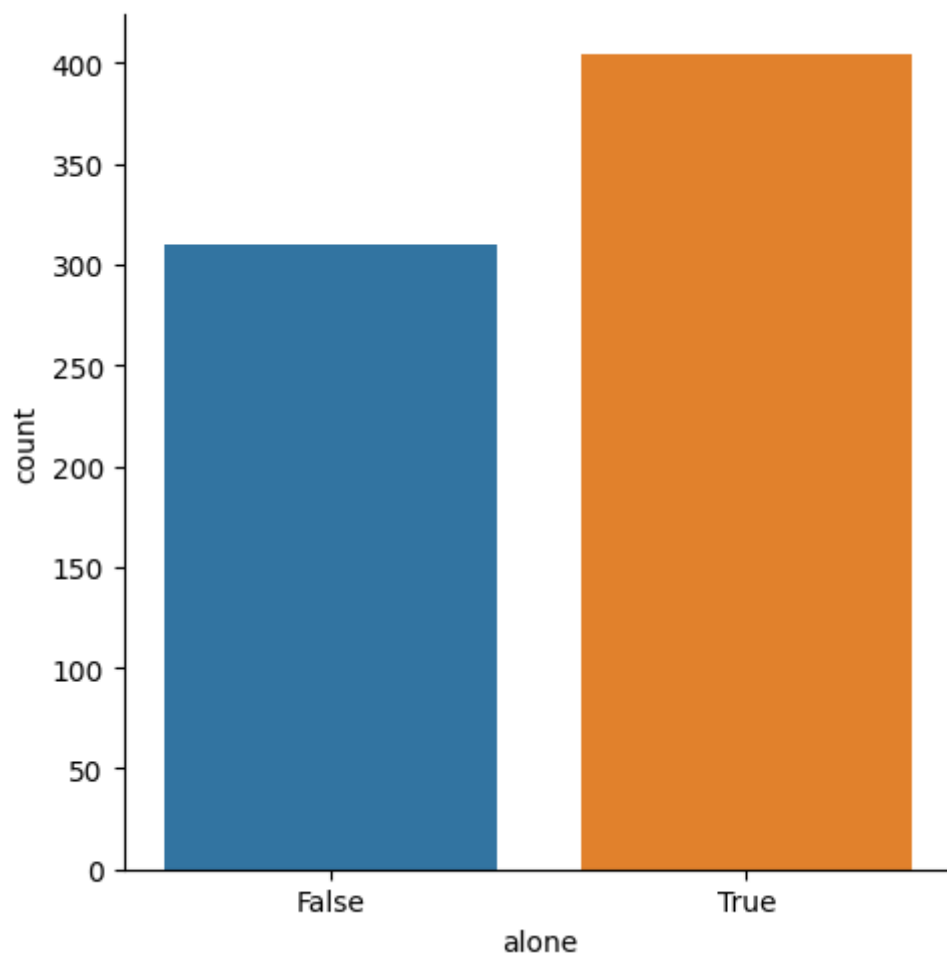
In [56]:



```
sns.catplot(x='alone',kind='count',data=df)
```

Out[56]:

```
<seaborn.axisgrid.FacetGrid at 0x207d6915c70>
```



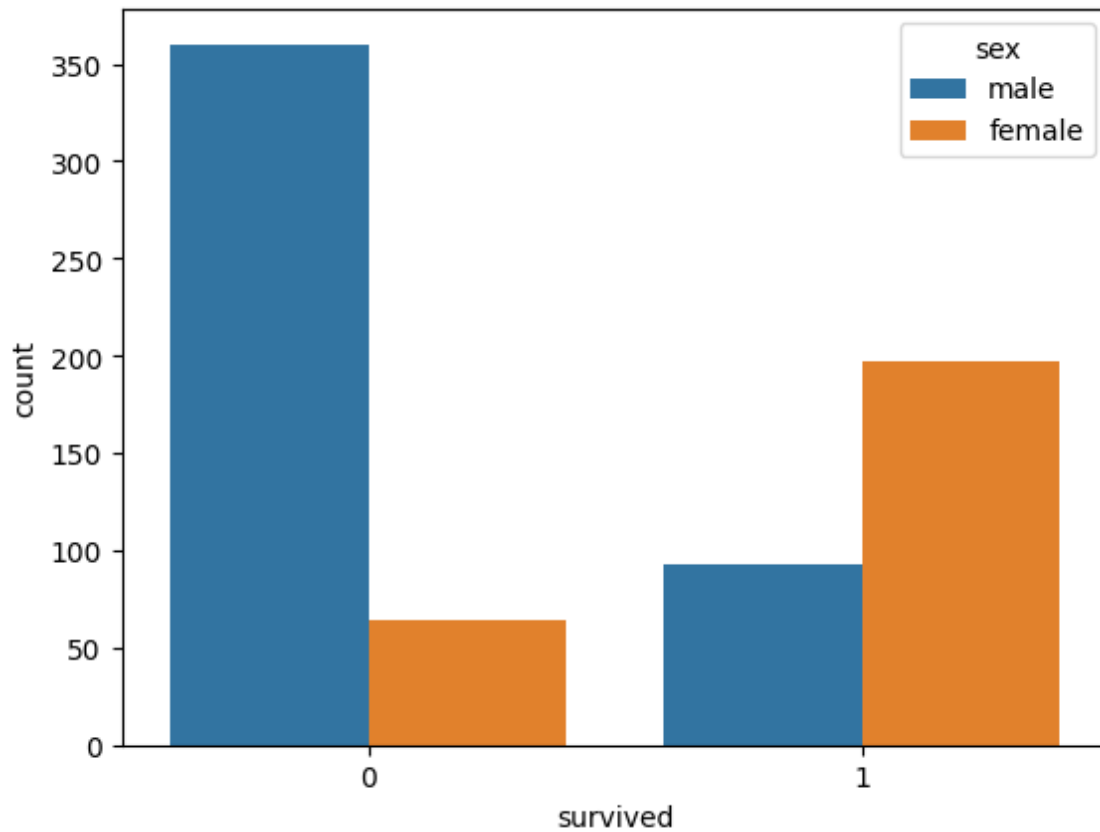
In [59]:



```
sns.countplot(x='survived',hue ='sex',data = df)
```

Out[59]:

```
<Axes: xlabel='survived', ylabel='count'>
```



In []:

