In [ ]:
```
#10. Data Visualization II
#1. Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box
#each gender along with the information about whether they survived or not. (C
#Write observations on the inference from the abovestatistics.

#dataset: titanic.csv
```

In [2]:
```
#titanic.csv
```

In [9]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df1=pd.read_csv("titanic.csv")
```

In [10]:
```
df1
```

Out[10]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	en
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	
...	...	...	...	...	...	...	...	...	...	...	
1305	3.0	0.0	Zabour, Miss. Thamine	female	NaN	1.0	0.0	2665	14.4542	NaN	
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5000	0.0	0.0	2656	7.2250	NaN	
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0000	0.0	0.0	2670	7.2250	NaN	
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0000	0.0	0.0	315082	7.8750	NaN	
1309	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

1310 rows × 14 columns

In [11]:
```
df=pd.DataFrame(df1)
df.head()
```

Out[11]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarke
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	

In [12]:
```
df.describe()
```

Out[12]:

	pclass	survived	age	sibsp	parch	fare	boc
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	121.00000
mean	2.294882	0.381971	29.881135	0.498854	0.385027	33.295479	160.80991
std	0.837836	0.486055	14.413500	1.041658	0.865560	51.758668	97.69692
min	1.000000	0.000000	0.166700	0.000000	0.000000	0.000000	1.00000
25%	2.000000	0.000000	21.000000	0.000000	0.000000	7.895800	72.00000
50%	3.000000	0.000000	28.000000	0.000000	0.000000	14.454200	155.00000
75%	3.000000	1.000000	39.000000	1.000000	0.000000	31.275000	256.00000
max	3.000000	1.000000	80.000000	8.000000	9.000000	512.329200	328.00000

In [13]:
```
df.info()
```

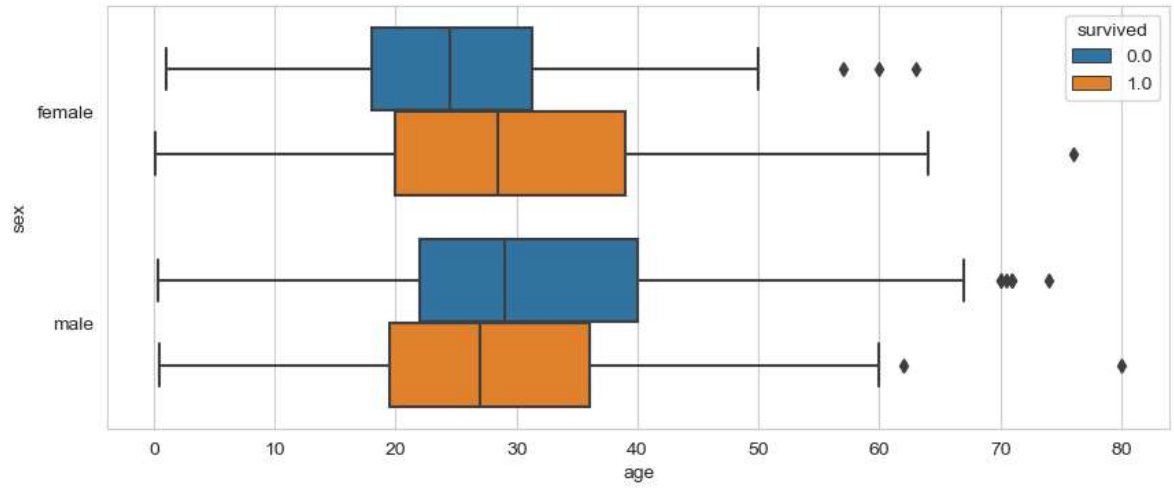
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1310 entries, 0 to 1309
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   pclass     1309 non-null   float64
 1   survived   1309 non-null   float64
 2   name       1309 non-null   object 
 3   sex        1309 non-null   object 
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   float64
 6   parch      1309 non-null   float64
 7   ticket     1309 non-null   object 
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object 
 10  embarked   1307 non-null   object 
 11  boat       486 non-null    object 
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object 
dtypes: float64(7), object(7)
memory usage: 143.4+ KB
```

In [14]:
```
df.columns
```

Out[14]:
```
Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticke
t',
       'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

In [15]:
```
sns.set_style('whitegrid')
plt.figure(figsize=(10,4))
sns.boxplot(x='age',y='sex',
            data=df,hue='survived')
```

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



In [ ]: