In [2]:
```
# 9) Data Visualization II
# 1. Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box plot for distribution of
# age with respect to each gender along with the information about whether they survived or
# not. (Column names : 'sex' and 'age')
# 2. Write observations on the inference from the above statistics.
```

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

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

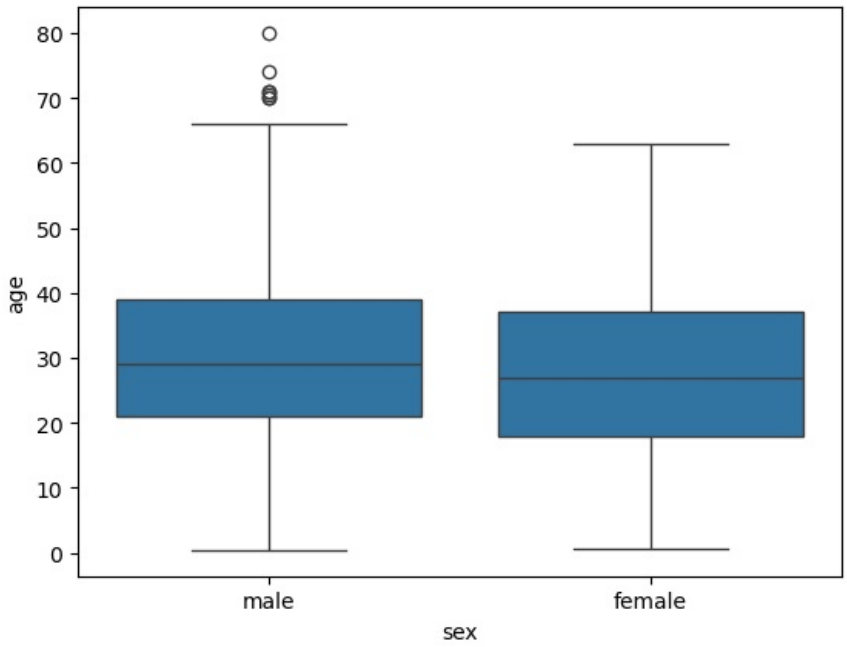
In [3]:
```
ds.head()
```

Out[3]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

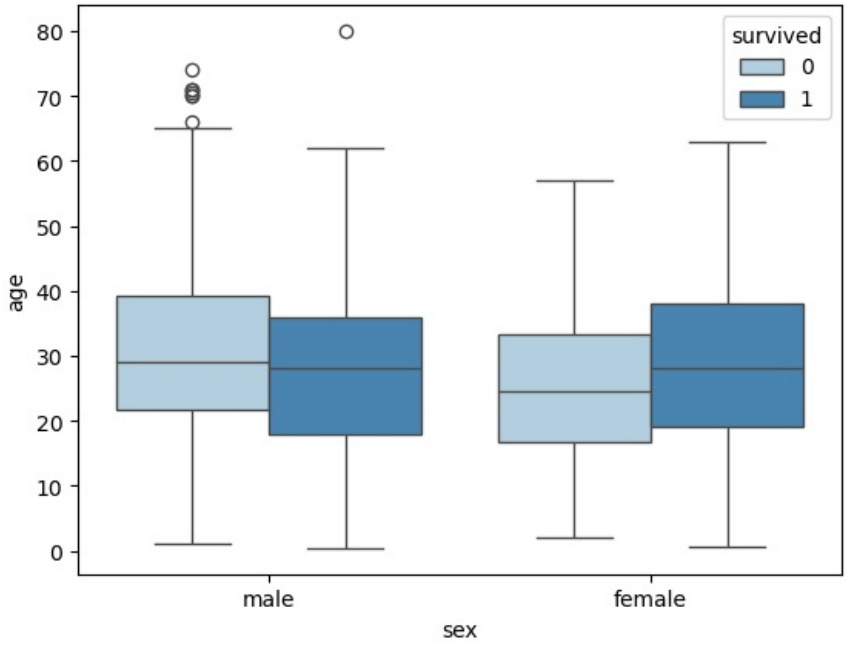
In [7]:
```
# (Boxplot Gender vs Age)
```

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



In [11]:
```
# Survived Passengers
```

In [19]:
```
sns.boxplot(x='sex', y='age', data=ds, hue='survived',  palette="Blues")
plt.show()
```



The first quartile starts at around 5 and ends at 22 which means that 25% of the passengers are aged between 5 and 25.

The second quartile starts at around 23 and ends at around 32 which means that 25% of the passengers are aged between 23 and 32.

Similarly, the third quartile starts and ends between 34 and 42, hence 25% passengers are aged within this range and

finally the fourth or last quartile starts at 43 and ends around 65.

Outliers: Any data points that fall outside the whiskers (i.e., the 1.5x IQR from Q1 and Q3) would be considered outliers. You may notice that there are a few points that fall outside the typical age ranges, representing passengers who were either very young or older than the majority.

In [ ]: