Assignment 9 - Data Visualization II (Data Science)

Atharva Taras (TE A - 73)

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')

Dataset - Kaggle Titanic (https://www.kaggle.com/competitions/titanic/data?select=train.csv)

In [1]:

```
1  import pandas as pd
2  import seaborn as sns
3  import matplotlib.pyplot as plt
```

In [2]:

```
1  data = pd.read_csv('titanic_train.csv')
2  data.head()
```
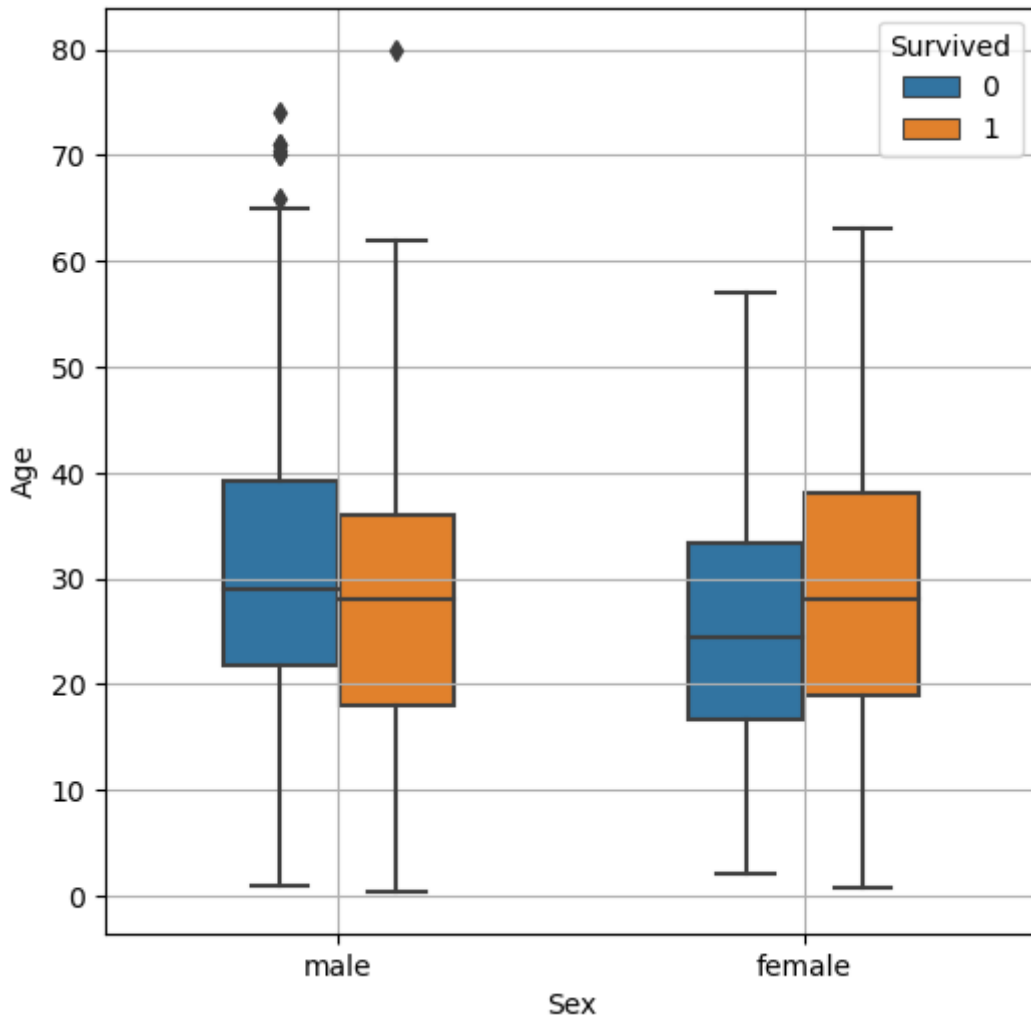
Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

```
1  plt.figure(figsize=(6, 6))
2  sns.boxplot(data=data, y='Age', x='Sex', hue='Survived', orient='v', width=0.5)
3  plt.grid()
```


2. Write observations on the inference from the above statistics

Observations
1 - Majority of male and female survivors are from the age group of 20-40
2 - Percenrage of non survivors is higher in case of males compared to females