

Practical No: A-9

- 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.

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
import seaborn as sns
df=sns.load_dataset("titanic")
df
```

↗

	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
...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	no	True
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes	True
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	no	False
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes	True
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	no	True

891 rows × 15 columns

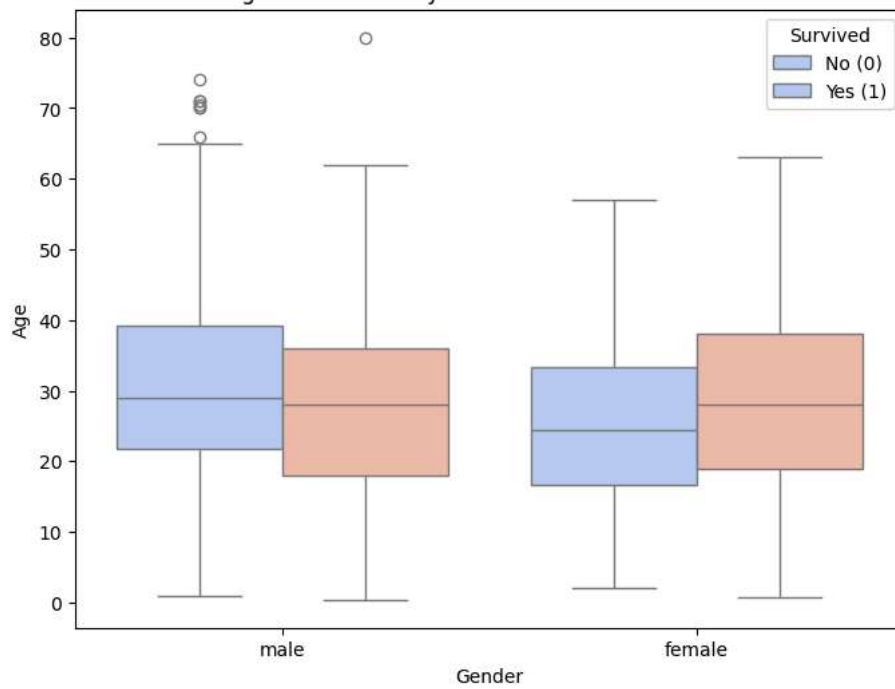
```
# Create the box plot
plt.figure(figsize=(8, 6))
sns.boxplot(x="sex", y="age", hue="survived", data=df, palette="coolwarm")

# Labels and title
plt.xlabel("Gender")
plt.ylabel("Age")
plt.title("Age Distribution by Gender and Survival Status")
plt.legend(title="Survived", labels=["No (0)", "Yes (1)"])

# Show the plot
plt.show()
```



Age Distribution by Gender and Survival Status



Observations from the Boxplot: Age Distribution Across Genders: The median age for both males and females is around 30–35 years. The interquartile range (IQR) (i.e., the middle 50% of the data) for both genders is quite similar, ranging approximately from 20 to 40 years. Survival vs. Non-Survival (Color-coded data): Males: The median age of non-survivors (blue) is slightly higher than survivors (orange). More outliers (older individuals) are present in the non-survivor category. Females: Survivors generally belong to a younger age group. Children and Outliers: Some outliers (marked as circles) are visible, particularly among older males who did not survive (age 70–80). The lower whisker extends close to 0 age, indicating the presence of children among both survivors and non-survivors. Key Inferences: