Practical 9

Title: Data Visualization III

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’)

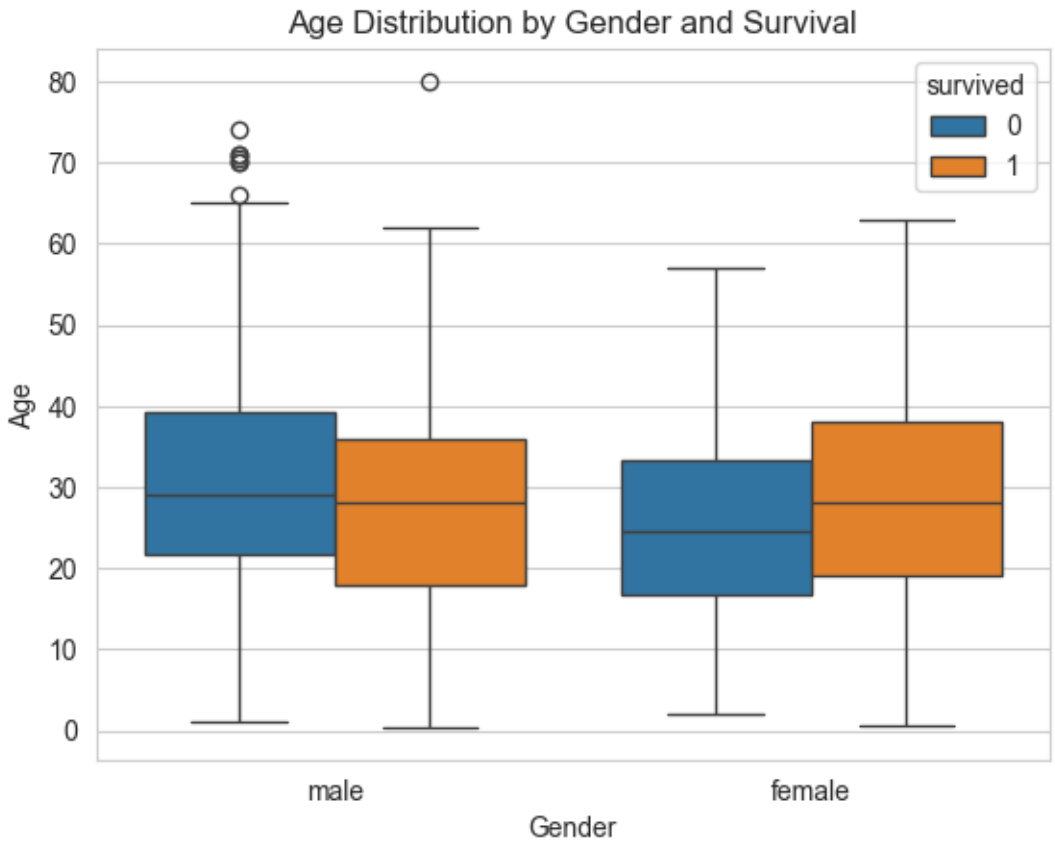
[ ]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[ ]:
```
# Load the Titanic dataset
titanic = sns.load_dataset('titanic')
```

[ ]:
```
# Set the plotting style
sns.set_style('whitegrid')
```

[ ]:
```
# Plot the box plot
sns.boxplot(x='sex', y='age', hue='survived', data=titanic)
# Set the title and labels
plt.title('Age Distribution by Gender and Survival')
plt.xlabel('Gender')
plt.ylabel('Age')

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



2. Write observations on the inference from the above statistics.

Observations from the statistics:

1. The box plot visually represents the distribution of ages for each gender and survival category. The boxes represent the interquartile range (IQR), the thick line inside the box represents the median, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR.
2. It can be observed that the median age for both males and females is relatively similar, with females tending to have a slightly higher median age.
3. The box plot shows that the age range for females who survived is wider compared to females who did not survive. This indicates that there were both younger and older females who survived the Titanic disaster.
4. For males, the box plot shows that the age range for those who survived is narrower compared to males who did not survive. This suggests that a larger proportion of younger males did not survive.
5. The box plot also shows the presence of outliers, represented as individual data points beyond the whiskers. These outliers might indicate unusual or extreme age values for certain individuals.