Data Visualization (TC-6)

LAB 3

> **Name: - Utkarsh Bhangale**
>
> **PRN: - 20200802124**

DS-1

> # Problem Statement

1. Data Visualization I

   - Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.
   - Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

2. Data Visualization II

   - 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')
   - Write observations on the inference from the above statistics.

> ## DV II

In [1]:

```python
# Importing the required libraries
import pandas as pd
import seaborn as sns
```

In [2]:

```python
df = pd.read_csv('/content/Titanic.csv')
```

In [3]:

```python
df.head(5)
```

Out[3]:

	sex	age	sibsp	parch	fare	embarked	class	who	alone	survived
0	male	22.0	1	0	7.2500	S	Third	man	False	0
1	female	38.0	1	0	71.2833	C	First	woman	False	1

	sex	age	sibsp	parch	fare	embarked	class	who	alone	survived
2	female	26.0	0	0	7.9250	S	Third	woman	True	1
3	female	35.0	1	0	53.1000	S	First	woman	False	1
4	male	35.0	0	0	8.0500	S	Third	man	True	0

In [4]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 10 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   sex       891 non-null    object
 1   age       714 non-null    float64
 2   sibsp     891 non-null    int64
 3   parch     891 non-null    int64
 4   fare      891 non-null    float64
 5   embarked  889 non-null    object
 6   class     891 non-null    object
 7   who       891 non-null    object
 8   alone     891 non-null    bool
 9   survived  891 non-null    int64
dtypes: bool(1), float64(2), int64(3), object(4)
memory usage: 63.6+ KB
```

In [6]:

```python
df.describe()
```

Out[6]:

	age	sibsp	parch	fare	survived
count	714.000000	891.000000	891.000000	891.000000	891.000000
mean	29.699118	0.523008	0.381594	32.204208	0.383838
std	14.526497	1.102743	0.806057	49.693429	0.486592
min	0.420000	0.000000	0.000000	0.000000	0.000000
25%	20.125000	0.000000	0.000000	7.910400	0.000000
50%	28.000000	0.000000	0.000000	14.454200	0.000000
75%	38.000000	1.000000	0.000000	31.000000	1.000000
max	80.000000	8.000000	6.000000	512.329200	1.000000

In [8]:
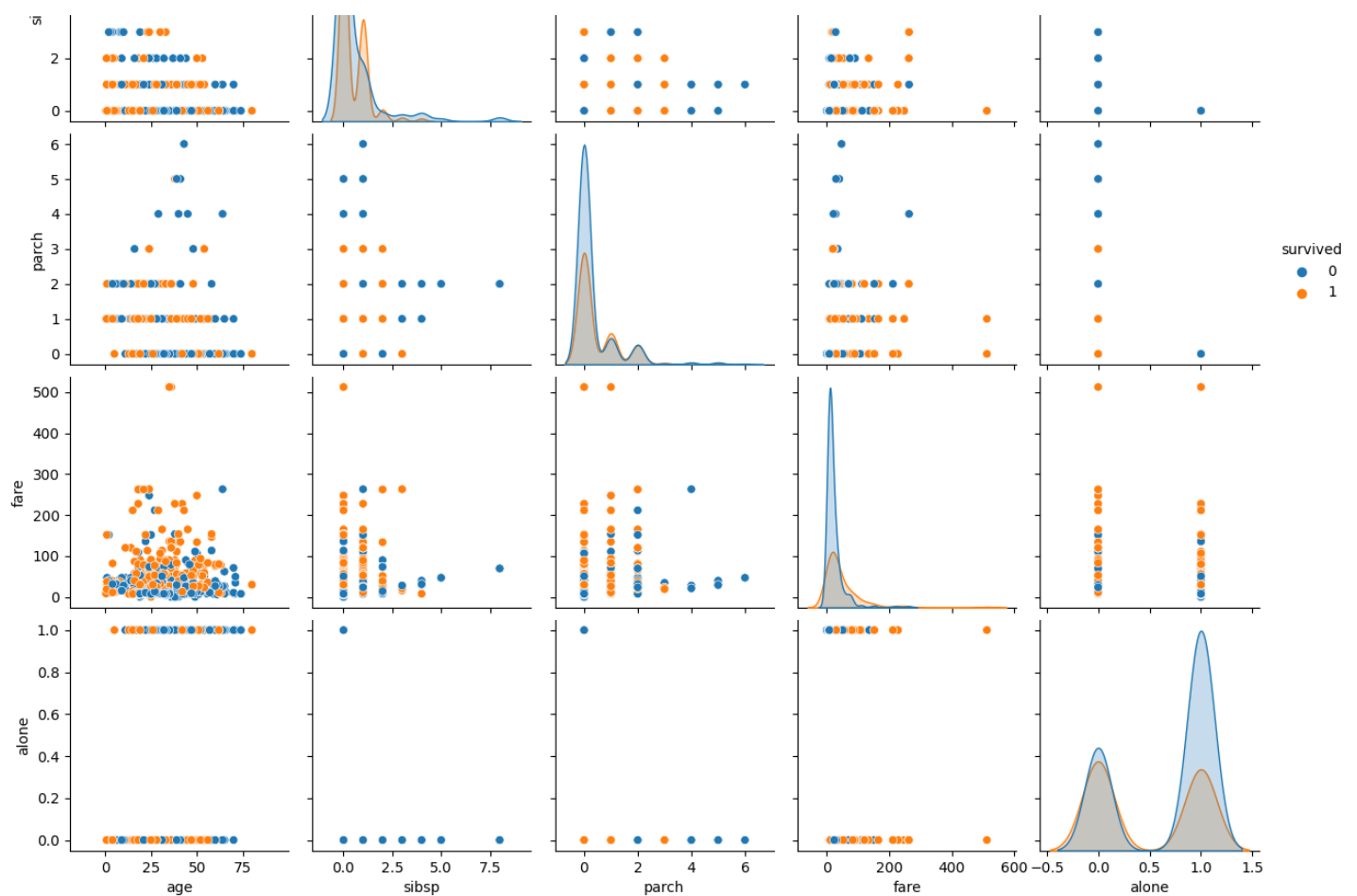
```python
# create a pairplot of the titanic dataset

sns.pairplot(df, hue='survived')
```

Out[8]:

```
<seaborn.axisgrid.PairGrid at 0x7e29f384e2c0>
```
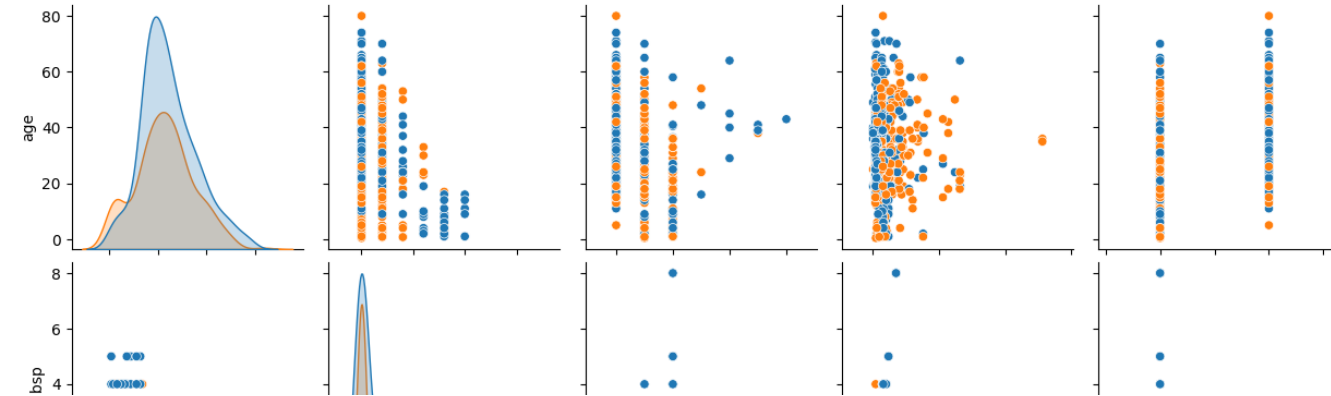
The pairplot shows the relationship between each pair of variables in the dataset. We can see some interesting patterns, such as:
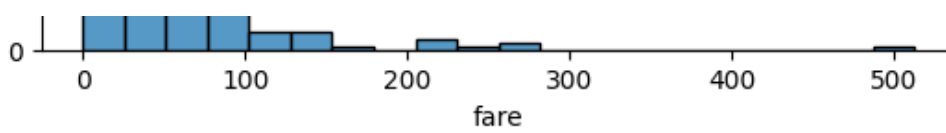
- Passengers who survived were more likely to be female and travel in first class.
- Passengers who traveled with siblings or parents were more likely to survive.
- Passengers who paid a higher fare were more likely to survive

---

In [9]:

```python
# Plot a histogram of the fare column
sns.histplot(data=df, x='fare', bins=20)
```

Out[9]:

```
<Axes: xlabel='fare', ylabel='Count'>
```
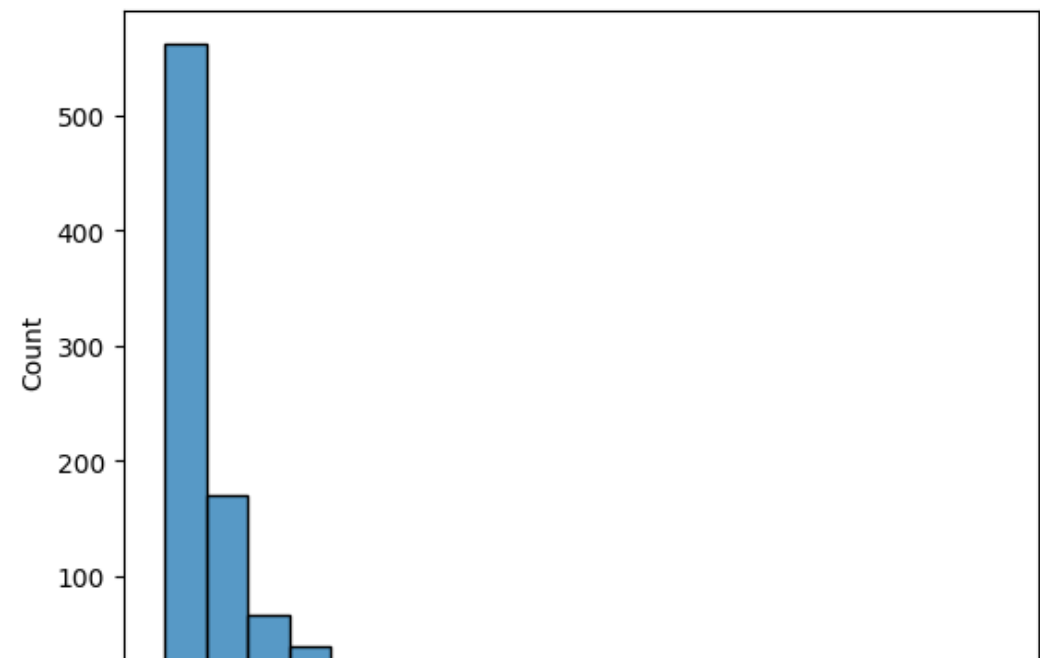
The histogram shows that the distribution of fares is right-skewed, meaning that there are more passengers who paid a lower fare than those who paid a higher fare.The most common fare is around $ 7.90, and the highest fare is

$512.33.

Overall, the pairplot and histogram show that there are some interesting patterns in the Titanic dataset. Passengers who survived were more likely to be female, travel in first class, travel with siblings or parents, and pay a higher fare
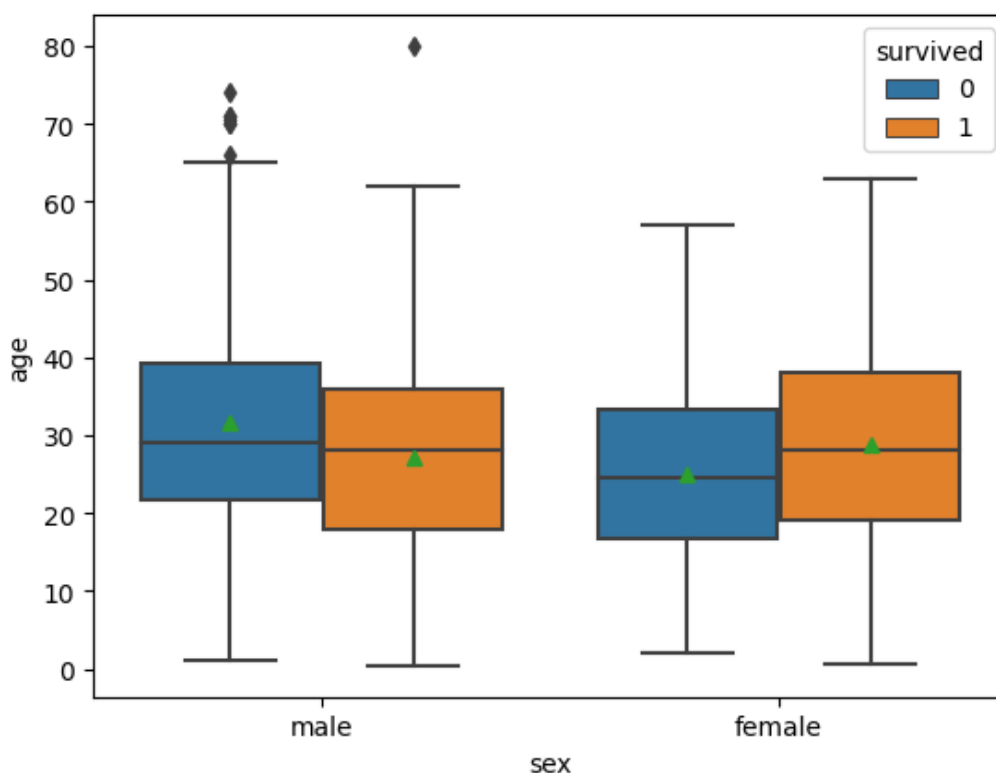
DV II

In [13]:

```python
# Loading the titanic dataset
titanic = sns.load_dataset("titanic")
```

In [14]:

```python
# Create a boxplot of age by gender and survival status.
sns.boxplot(
    x="sex",
    y="age",
    hue="survived",
    showmeans=True,
    data=titanic,
)
```

Out[14]:

```
<Axes: xlabel='sex', ylabel='age'>
```

Observations:

- On the inference from the above statistics: The boxplot shows that female passengers were generally younger than male passengers, and that passengers who survived were generally younger than passengers who did not survive.
- The median age for female passengers who survived is around 26 years old, while the median age for female passengers who did not survive is around 30 years old. The median age for male passengers who survived is around 28 years old, while the median age for male passengers who did not survive is around 31 years old.
- The boxplot also shows that there is a wider range of ages for male passengers who did not survive than for female passengers who did not survive. This suggests that there may have been more factors that contributed to the deaths of male passengers than female passengers.

Conclusion :

Overall, the boxplot provides some evidence that age was a factor in survival on the Titanic.Passengers who were younger were more likely to survive than passengers who were older. However, it is important to note that age was just one factor that contributed to survival. Other factors, such as social class and whether or not the passenger was traveling with family, also played a role.