

In [2]: `pip install seaborn`

```
Collecting seaborn
  Downloading seaborn-0.11.2-py3-none-any.whl (292 kB)
    292.8/292.8 KB 1.7 MB/s eta 0:00:00[36m
0:00:01m eta 0:00:01
Requirement already satisfied: numpy>=1.15 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from seaborn) (1.22.3)
Collecting scipy>=1.0
  Downloading scipy-1.8.0-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (4
1.6 MB)
    41.6/41.6 MB 3.0 MB/s eta 0:00:00m eta
0:00:01[36m0:00:01
Requirement already satisfied: pandas>=0.23 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from seaborn) (1.4.2)
Requirement already satisfied: matplotlib>=2.2 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from seaborn) (3.5.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (1.4.2)
Requirement already satisfied: fonttools>=4.22.0 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (4.33.3)
Requirement already satisfied: pillow>=6.2.0 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (9.1.0)
Requirement already satisfied: pyparsing>=2.2.1 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (3.0.7)
Requirement already satisfied: python-dateutil>=2.7 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (2.8.2)
Requirement already satisfied: packaging>=20.0 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (21.3)
Requirement already satisfied: cycler>=0.10 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from matplotlib>=2.2->seaborn) (0.11.0)
Requirement already satisfied: pytz>=2020.1 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from pandas>=0.23->seaborn) (2022.1)
Requirement already satisfied: six>=1.5 in /home/csl2/notebook/jupyterenv/lib/python3.8/site-packages (from python-dateutil>=2.7->matplotlib>=2.2->seaborn) (1.16.0)
Installing collected packages: scipy, seaborn
Successfully installed scipy-1.8.0 seaborn-0.11.2
Note: you may need to restart the kernel to use updated packages.
```

In [3]: `import pandas as pd`  
`import numpy as np`  
  
`import matplotlib.pyplot as plt`  
`import seaborn as sns`  
  
`dataset = sns.load_dataset('titanic')`  
  
`dataset.head()`

Matplotlib is building the font cache; this may take a moment.

```
Out[3]:
```

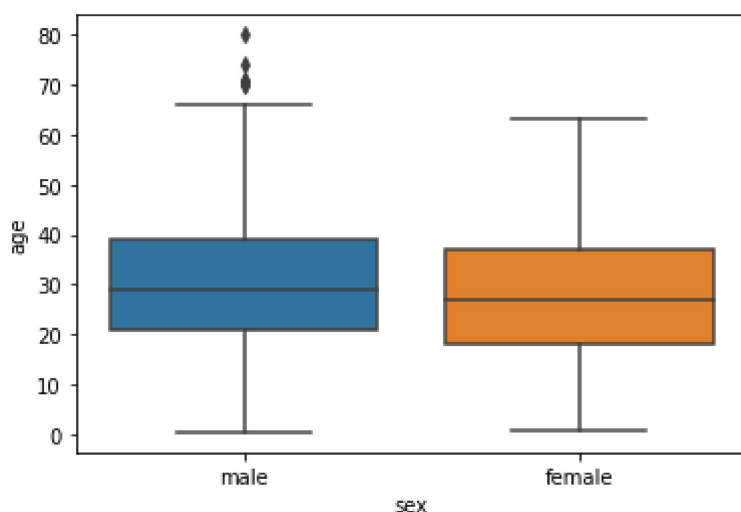
|   | survived | pclass | sex    | age  | sibsp | parch | fare    | embarked | class | who   | adult_male | deck |
|---|----------|--------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|
| 0 | 0        | 3      | male   | 22.0 | 1     | 0     | 7.2500  | S        | Third | man   | True       | NaN  |
| 1 | 1        | 1      | female | 38.0 | 1     | 0     | 71.2833 | C        | First | woman | False      | C    |
| 2 | 1        | 3      | female | 26.0 | 0     | 0     | 7.9250  | S        | Third | woman | False      | NaN  |
| 3 | 1        | 1      | female | 35.0 | 1     | 0     | 53.1000 | S        | First | woman | False      | C    |
| 4 | 0        | 3      | male   | 35.0 | 0     | 0     | 8.0500  | S        | Third | man   | True       | NaN  |

```
In [4]: # Box Plot
```

```
In [5]: #Now Let's plot a box plot that displays the distribution for the age with respect to
#You need to pass the categorical column as the first parameter (which is sex in our case)
#and the numeric column (age in our case) as the second parameter.
#Finally, the dataset is passed as the third parameter
```

```
In [6]: sns.boxplot(x='sex', y='age', data=dataset)
```

```
Out[6]: <AxesSubplot:xlabel='sex', ylabel='age'>
```



```
In [7]: #Let's try to understand the box plot for female.
#The first quartile starts at around 5 and ends at 22 which means that 25% of the passengers
#The second quartile starts at around 23 and ends at around 32 which means that 25% of the passengers
#Similarly, the third quartile starts and ends between 34 and 42, hence 25% passengers
#the fourth or last quartile starts at 43 and ends around 65.
```

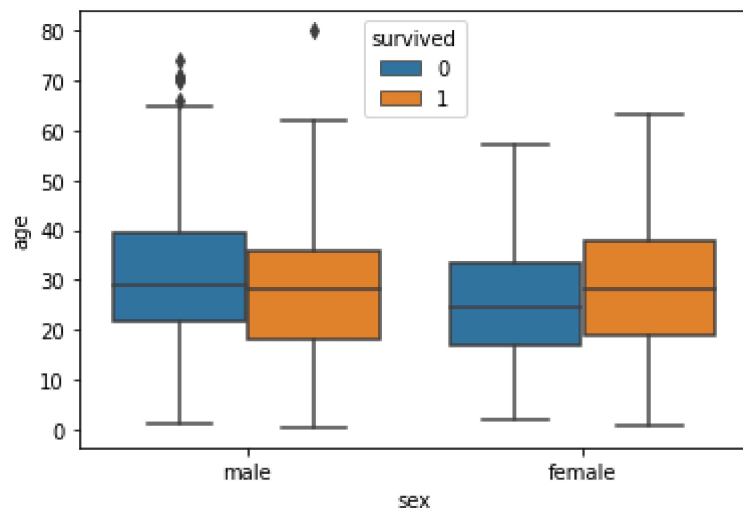
```
#If there are any outliers or the passengers that do not belong to any of the quartile
#they are called outliers and are represented by dots on the box plot.
```

```
In [ ]:
```

```
In [8]: #You can make your box plots more fancy by adding another layer of distribution.
#For instance, if you want to see the box plots of forage of passengers of both gender
#along with the information about whether or not they survived, you can pass the survived
#as value to the hue parameter as shown below:
```

```
In [9]: sns.boxplot(x='sex', y='age', data=dataset, hue="survived")
```

Out[9]: <AxesSubplot:xlabel='sex', ylabel='age'>



In [ ]: