Experiment-4

4) Consider auto-mpg dataset and titanic dataset perform Univariate and Bivariate analysis plot the graph

Define Univariate

Univariate data refers to a dataset where each observation is associated

with only one variable.

This means it focuses on measuring or observing a single characteristic

or attribute for each individual in the dataset.

```python
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         df=pd.read_csv('C:/Users/Guru Kiran/All CSV files/auto-mpg.csv')
         df
```

|	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin
0	18.0	8	307.0	130	3504	12.0	70	1
1	15.0	8	350.0	165	3693	11.5	70	1
2	18.0	8	318.0	150	3436	11.0	70	1
3	16.0	8	304.0	150	3433	12.0	70	1
4	17.0	8	302.0	140	3449	10.5	70	1
...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1
394	44.0	4	97.0	52	2130	24.6	82	2
395	32.0	4	135.0	84	2295	11.6	82	1
396	28.0	4	120.0	79	2625	18.6	82	1
397	31.0	4	119.0	82	2720	19.4	82	1

398 rows × 9 columns

Using univariate ploting Histogram

In [2]:
```python
plt.hist(df['cylinders'])
plt.title('counting cylinders')
plt.xlabel('cylinders')
plt.ylabel('cylinders counts')
plt.show()
```
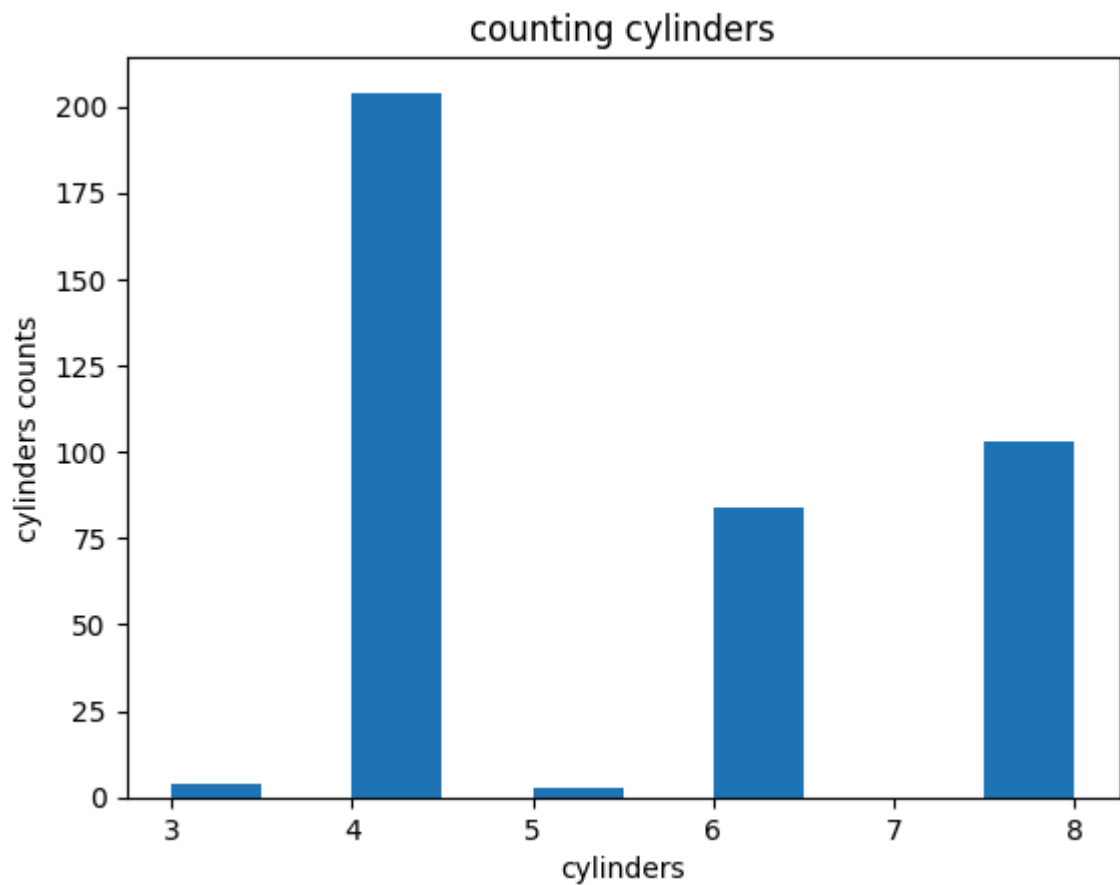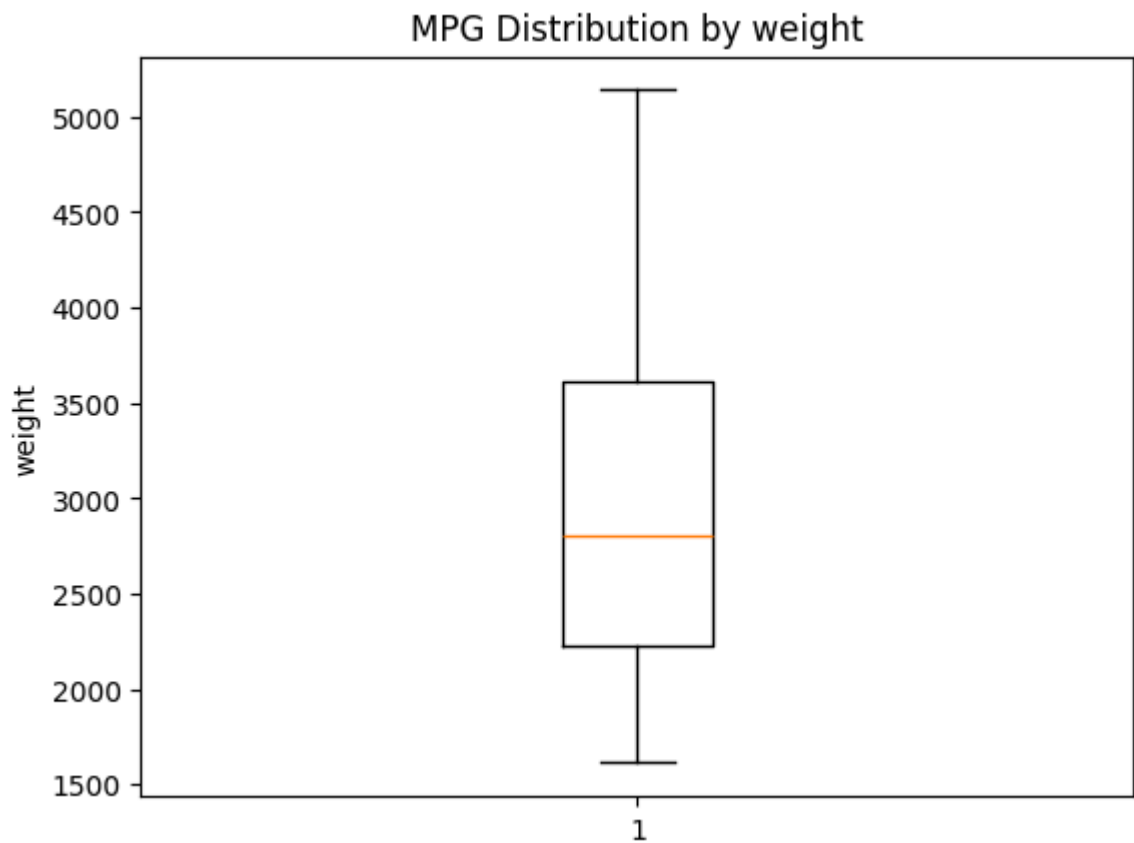
Using univariate ploting line graph

```python
In [3]: df['model year'].value_counts().sort_index().plot(kind='line')
        plt.xlabel('Model Year')
        plt.ylabel('Number of Cars')
        plt.title('Number of Cars by Model Year')
        plt.show()
```

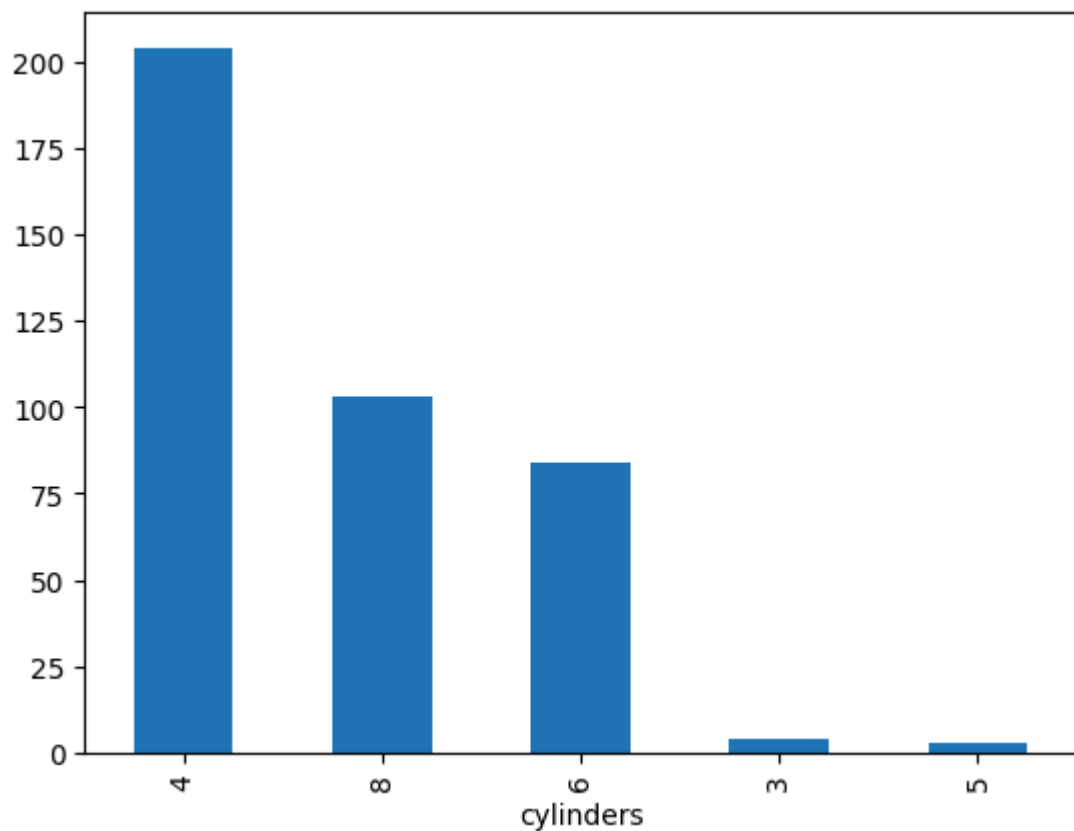Number of Cars by Model Year
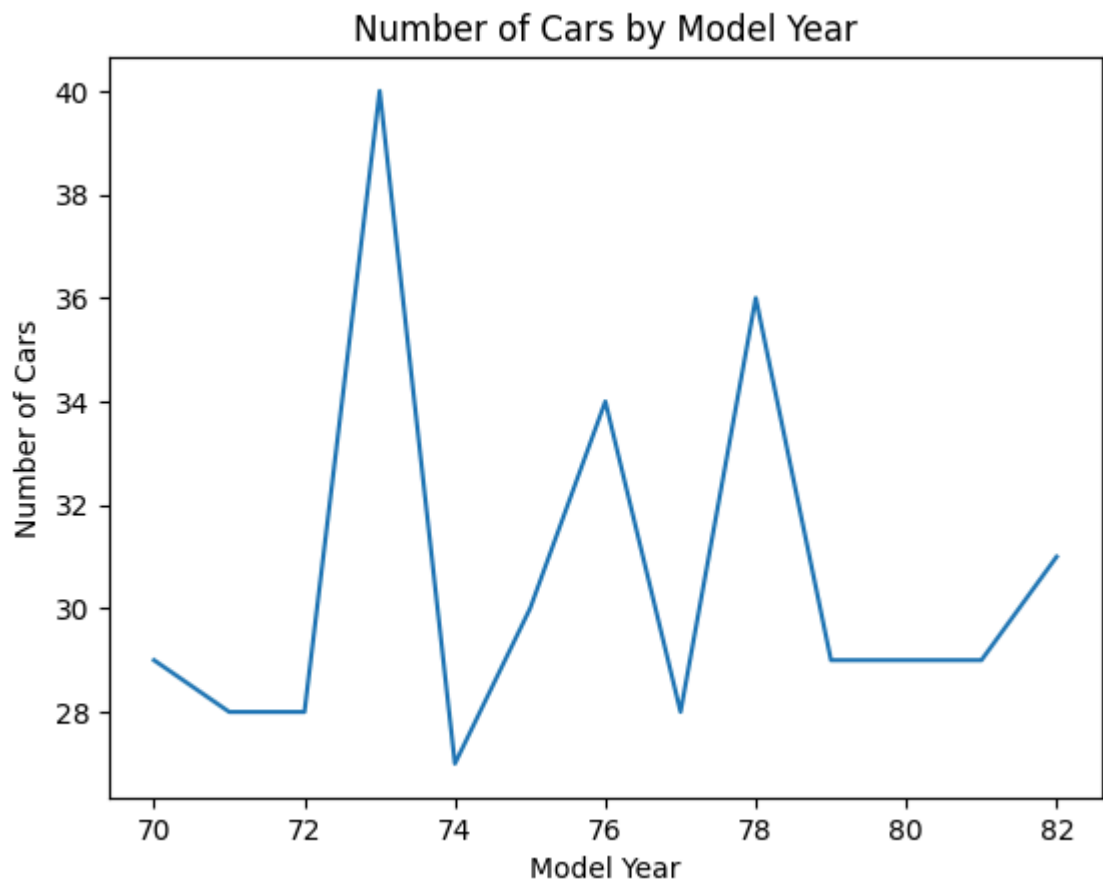
Using univariate ploting boxplot

```
In [4]:  plt.boxplot(df['weight'])
         plt.title('MPG Distribution by weight')
         plt.ylabel('weight')
         plt.show()
```

MPG Distribution by weight

Using univariate ploting bar graph

```
In [5]: df['cylinders'].value_counts().plot(kind='bar')
```
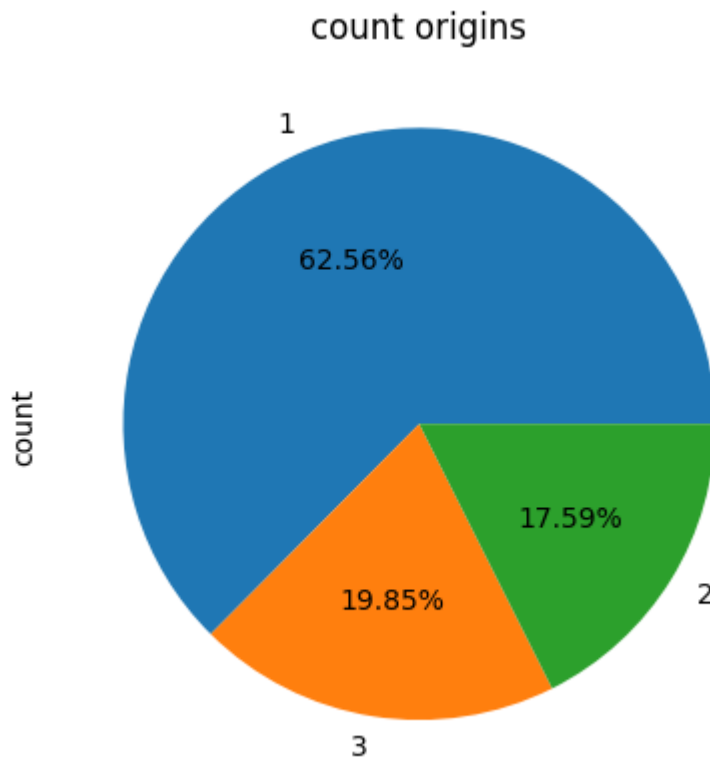
Out[5]: <Axes: xlabel='cylinders'>

Using univariate ploting pie chart

```
In [6]: df['origin'].value_counts().plot(kind='pie',autopct='%1.2f%%')
        plt.title('count origins')
```

```
Out[6]: Text(0.5, 1.0, 'count origins')
```


count origins

Bivariate Analysis

Using titanic dataset

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        df=pd.read_csv('C:/Users/Guru Kiran/All CSV files/titanic.csv')
        df.head(5)
```
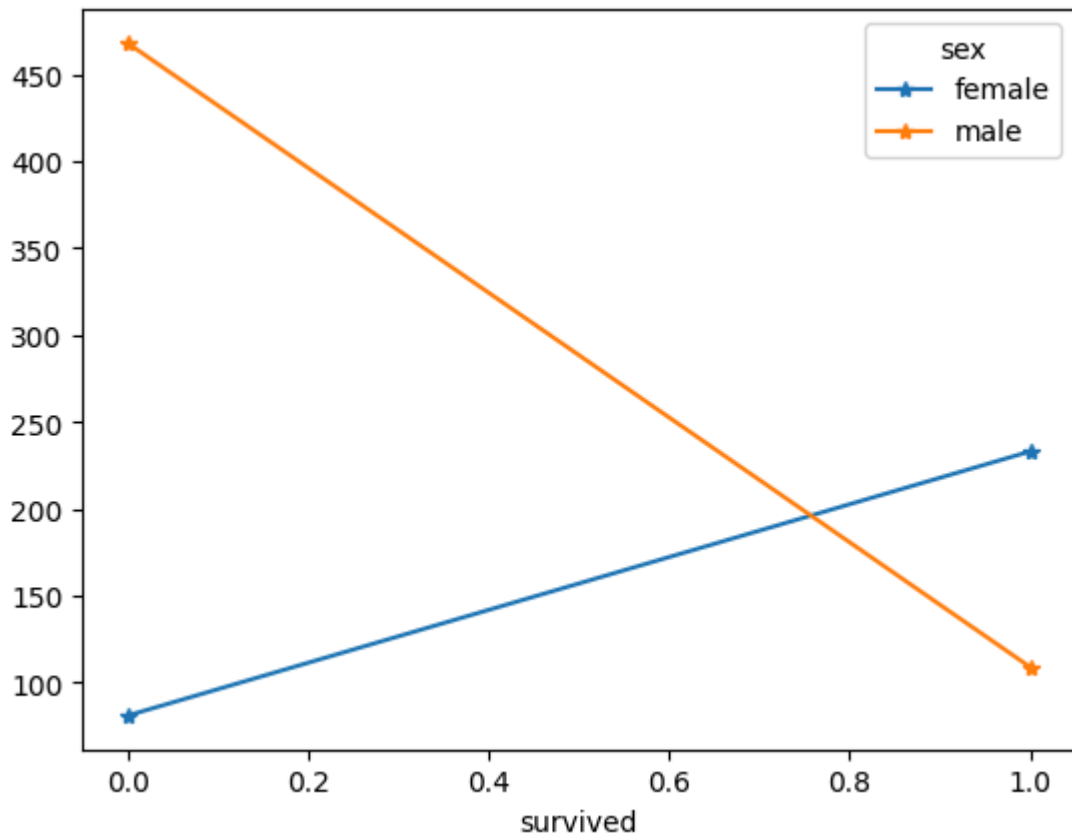
Out[1]:

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

Using bivariate ploting line graph

```
a=df.groupby(['survived','sex']).size().unstack()
a.plot(kind='line',marker='*')
```
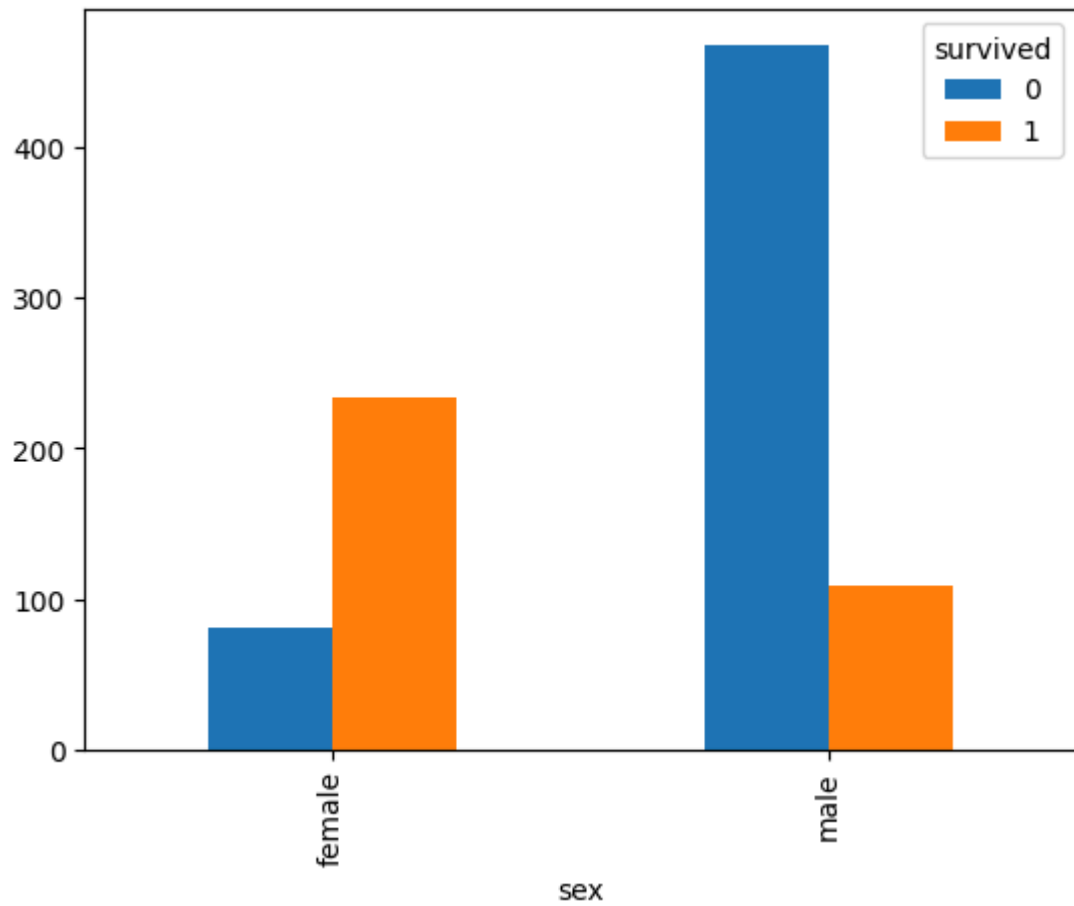
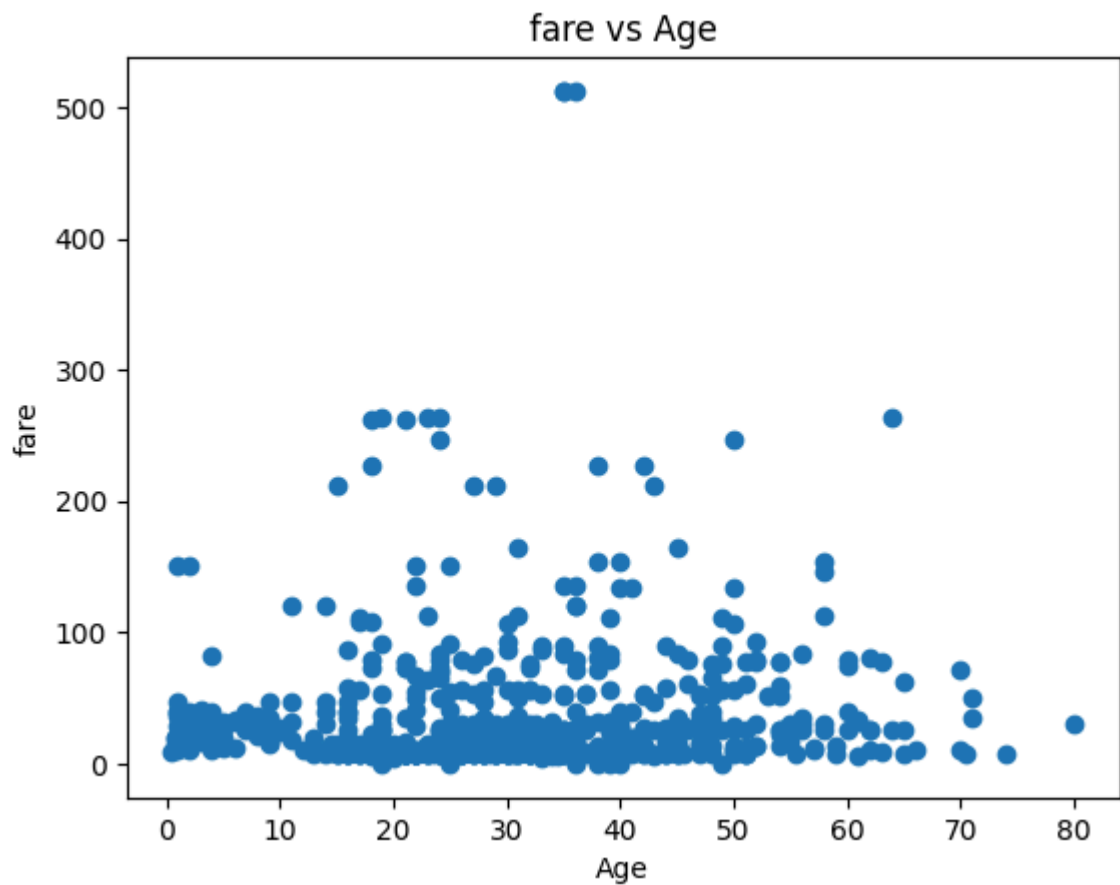`<Axes: xlabel='survived'>`

Using bivariate ploting bar plot

```
a=df.groupby(['sex','survived']).size().unstack()
a.plot(kind='bar')
```

`<Axes: xlabel='sex'>`

Using bivariate ploting scatter plot

```
In [10]: plt.scatter(df['age'], df['fare'])
         plt.xlabel('Age')
         plt.ylabel('fare')
         plt.title('fare vs Age')
         plt.show()
```

fare vs Age

Multivaria analysis

```
In [11]:  a=df.groupby(['sex','survived','who']).size().unstack()
          a.plot(kind='bar')
```

Out[11]: <Axes: xlabel='sex,survived'>