

```
In [78]: %cd C:\Users\manoj\Downloads
```

C:\Users\manoj\Downloads

```
In [79]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [80]: # naming of columns
df=pd.read_csv('haberman.csv',names=['A','Yoo','Positivenodes','Survivalrate'])
df
```

0	30	64	1	1
1	30	62	3	1
2	30	65	0	1
3	31	59	2	1
4	31	65	4	1
...	...	...	...	...
301	75	62	1	1
302	76	67	0	1
303	77	65	3	1
304	78	65	1	2
305	83	58	2	2

306 rows × 4 columns

```
In [81]: #renaming of column ,pass in dictionary  
df.rename({'A':'Age','Yoo':'Yearofoperation'},axis=1,inplace=True)  
df
```

Out[81]:

	Age	Yearofoperation	Positivenodes	Survivalrate
0	30	64	1	1
1	30	62	3	1
2	30	65	0	1
3	31	59	2	1
4	31	65	4	1
...	...	...	...	...
301	75	62	1	1
302	76	67	0	1
303	77	65	3	1
304	78	65	1	2
305	83	58	2	2

306 rows × 4 columns

```
In [82]: #information
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306 entries, 0 to 305
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Age              306 non-null   int64
1   Yearofoperation  306 non-null   int64
2   Positivenodes    306 non-null   int64
3   Survivalrate     306 non-null   int64
dtypes: int64(4)
memory usage: 9.7 KB
```

```
In [83]: #finding null values
df.isnull().sum()
```

```
Out[83]: Age              0
Yearofoperation          0
Positivenodes            0
Survivalrate             0
dtype: int64
```

```
In [84]: #type of datas
df['Survivalrate'].unique()
```

```
Out[84]: array([1, 2], dtype=int64)
```

```
In [85]: # counting data
df['Survivalrate'].value_counts()
k=np.array(df.Survivalrate)
count1=0
count2=0
for j in k:
    if j==1:
        count1=count1+1
    elif j==2:
        count2=count2+1
print('patients survived 5 years and longer are',count1,'patients died with in 5 years are',count2)
```

patients survived 5 years and longer are 225 patients died with in 5 years are 81

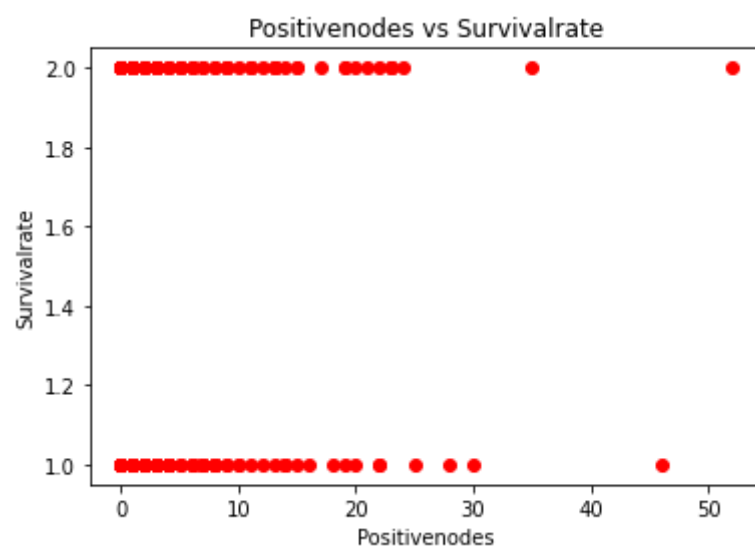
```
In [86]: #scatter plot
plt.scatter(df['Age'],df['Yearofoperation'])
plt.xlabel('Age')
plt.ylabel('Yearofoperation')
plt.title('Age vs Yearofoperation')
```

Out[86]: Text(0.5, 1.0, 'Age vs Yearofoperation')



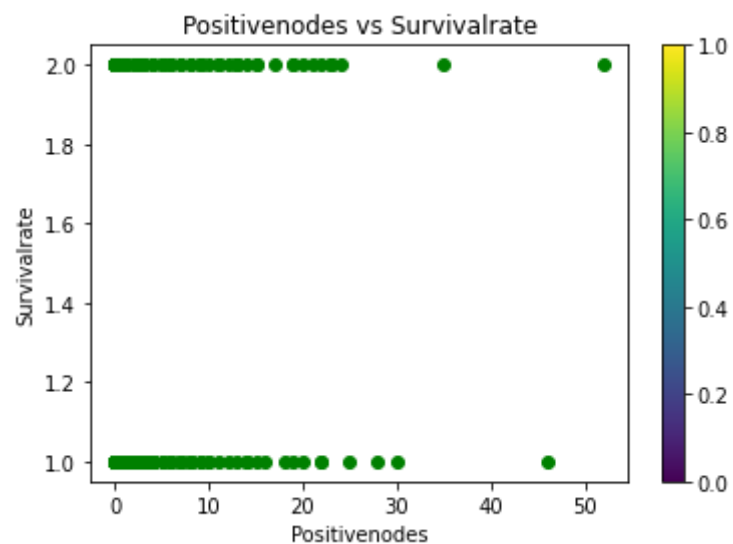
```
In [112]: #color change  
plt.scatter(df['Positivenodes'],df['Survivalrate'],c='red')  
plt.xlabel('Positivenodes')  
plt.ylabel('Survivalrate')  
plt.title('Positivenodes vs Survivalrate')
```

```
Out[112]: Text(0.5, 1.0, 'Positivenodes vs Survivalrate')
```



```
In [88]: #colorbar ,cmap
plt.scatter(df['Positivenodes'],df['Survivalrate'],c='green',cmap='flag')
plt.xlabel('Positivenodes')
plt.ylabel('Survivalrate')
plt.title('Positivenodes vs Survivalrate')
plt.colorbar()
```

Out[88]: <matplotlib.colorbar.Colorbar at 0x235c56fa160>

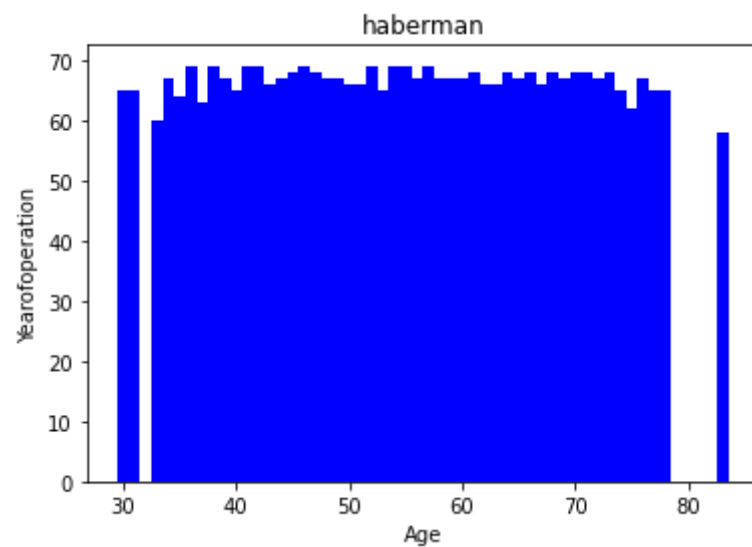


```
In [89]: #transparency
plt.scatter(df['Age'],df['Yearofoperation'],alpha=0.5)
plt.xlabel('Age')
plt.ylabel('Yearofoperation')
plt.title('Age vs Yearofoperation')
```

```
Out[89]: Text(0.5, 1.0, 'Age vs Yearofoperation')
```



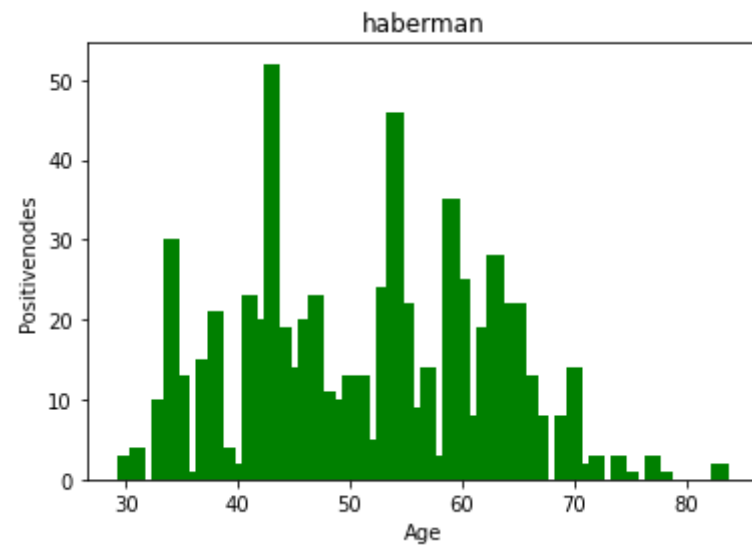
```
In [90]: df=pd.read_csv('haberman.csv',names=['Age','Yearofoperation','Positivenodes','Survivalrate'])
df
plt.bar(df['Age'],df['Yearofoperation'],color='b',width=1)
plt.title('haberman')
plt.xlabel('Age')
plt.ylabel('Yearofoperation')
plt.show()
```





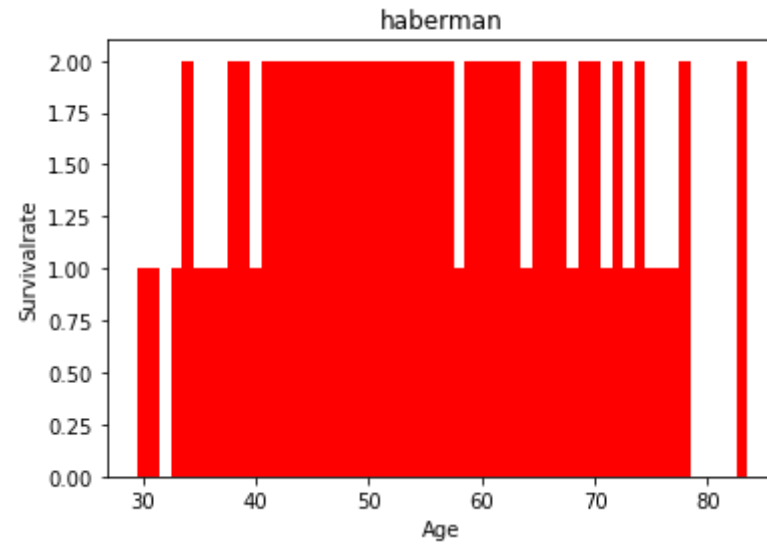
```
In [91]: plt.bar(df['Age'],df['Positivenodes'],color='green',width=1.5)
plt.title('haberman')
plt.xlabel('Age')
plt.ylabel('Positivenodes')
```

```
Out[91]: Text(0, 0.5, 'Positivenodes')
```



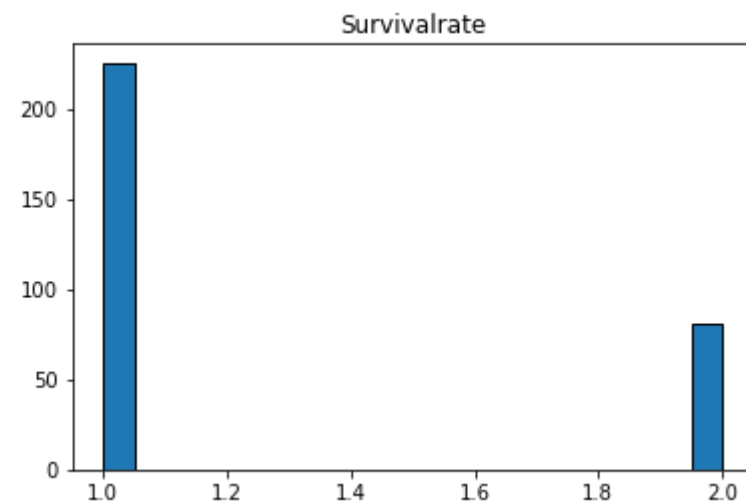
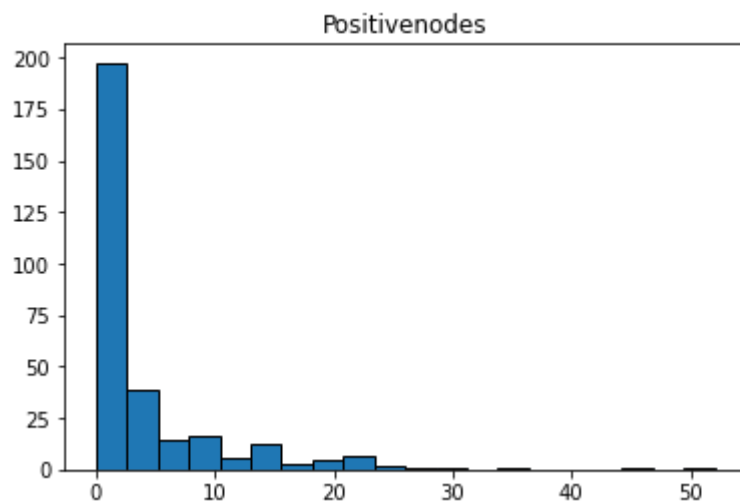
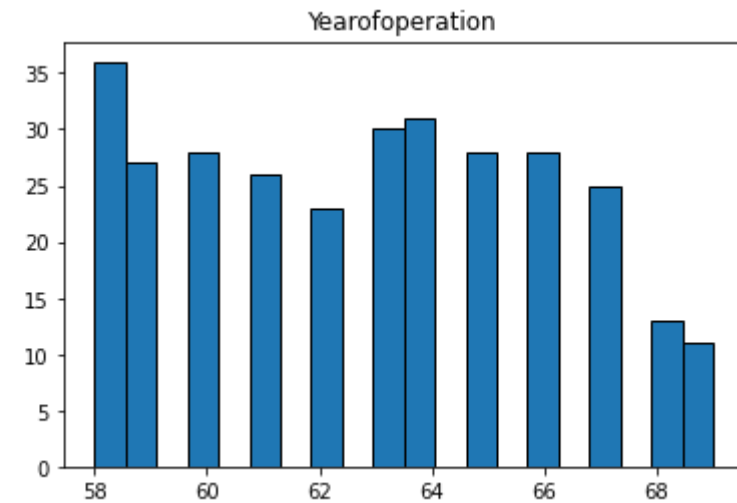
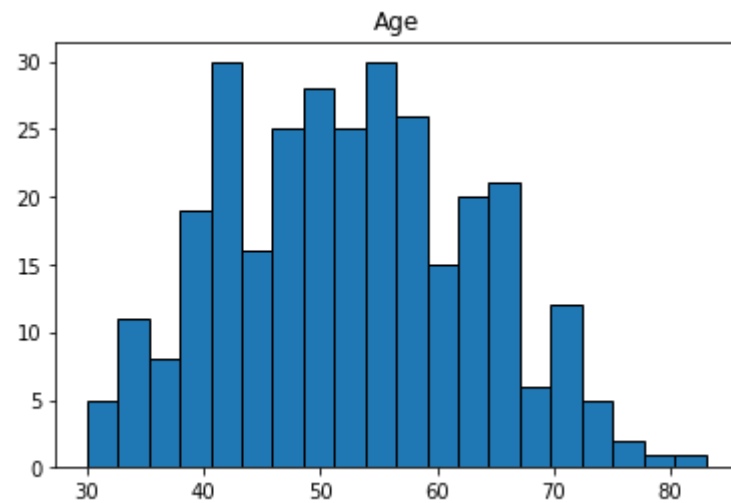
```
In [92]: plt.bar(df['Age'],df['Survivalrate'],color='red',width=1)  
plt.title('haberman')  
plt.xlabel('Age')  
plt.ylabel('Survivalrate')
```

```
Out[92]: Text(0, 0.5, 'Survivalrate')
```



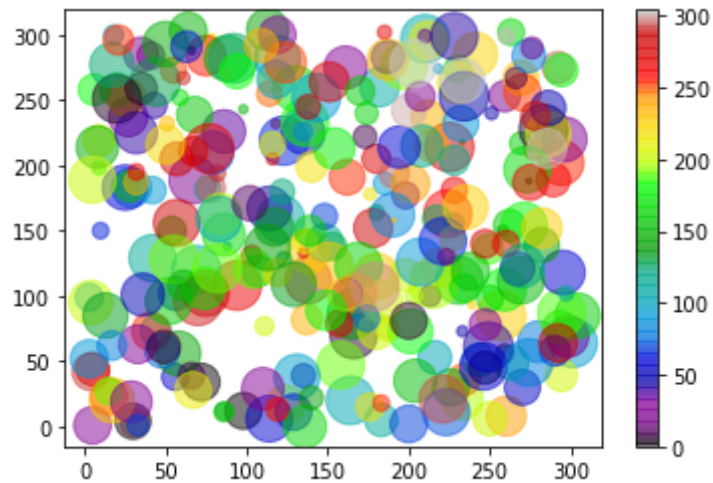
```
In [93]: df[['Age', 'Yearofoperation', 'Positivenodes', 'Survivalrate']].hist(figsize=(14,9),bins=20,linewidth='1',edgecolor='k')
```

```
Out[93]: array([[<AxesSubplot:title={'center':'Age'}>,  
  <AxesSubplot:title={'center':'Yearofoperation'}>],  
  [<AxesSubplot:title={'center':'Positivenodes'}>,  
  <AxesSubplot:title={'center':'Survivalrate'}>]], dtype=object)
```



```
In [94]: #combine color size and alpha
x=np.random.randint(305,size=(305))
y=np.random.randint(305,size=(305))
colors=np.random.randint(305,size=(305))
sizes=2*np.random.randint(305,size=(305))
plt.scatter(x,y,c=colors,s=sizes,alpha=0.5,cmap='nipy_spectral')
plt.colorbar()
```

Out[94]: <matplotlib.colorbar.Colorbar at 0x235c72e44f0>



```
In [108]: y=np.array(df["Age"])  
plt.legend(title="Ages")  
plt.pie(y)  
plt.show()
```

No handles with labels found to put in legend.

