In [3]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

In [2]:

```
import warnings
warnings.filterwarnings("ignore")
```

In [1]:

```
from google.colab import files
files=files.upload()
```

Choose File **No file selected**

**Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.**

```
Saving haberman (2).csv to haberman (2).csv
```

In [5]:

```
data=pd.read_csv('haberman (2).csv')
data.head()
```

Out[5]:

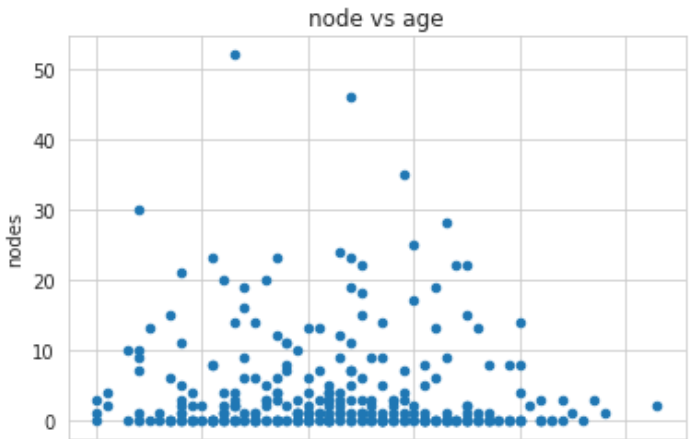
| | age | year | nodes | status |
|---|---|---|---|---|
| **0** | 30 | 64 | 1 | 1 |
| **1** | 30 | 62 | 3 | 1 |
| **2** | 30 | 65 | 0 | 1 |
| **3** | 31 | 59 | 2 | 1 |
| **4** | 31 | 65 | 4 | 1 |

**objective:To find if patients survive after 5 years or died with 5 year,after doing operation**

**SCATTER PLOT**

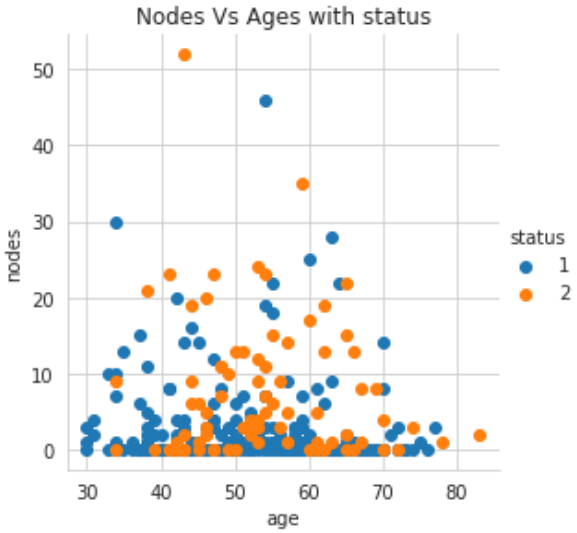
In [8]:

```
#2D SCATTER PLOT
data.plot(kind='scatter', x='age', y='nodes') ;
plt.title('node vs age')
plt.show()
```




In [9]:

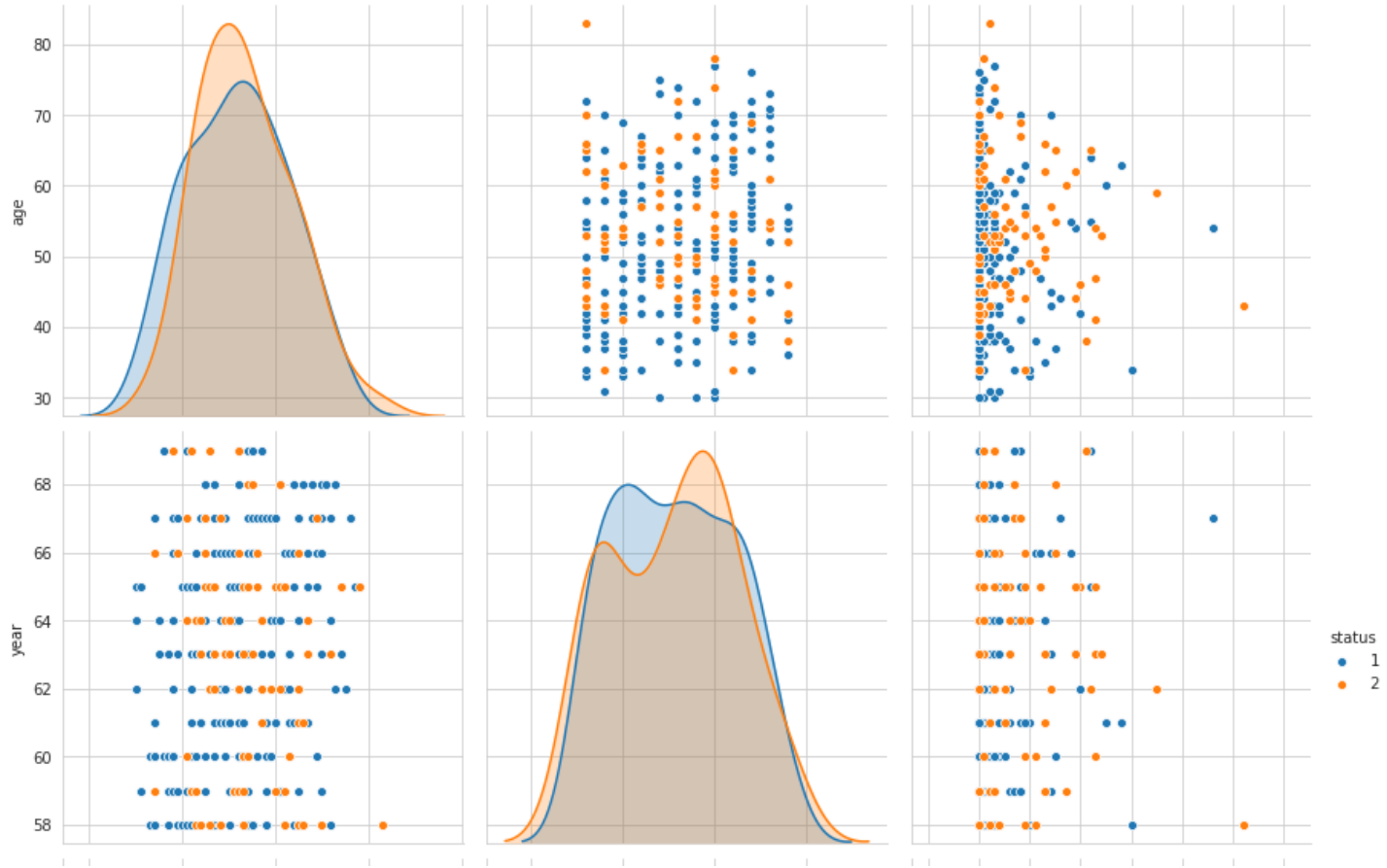
```
sns.set_style("whitegrid");
sns.FacetGrid(data, hue="status", size=4) \
   .map(plt.scatter, "age","nodes") \
   .add_legend();

plt.title('Nodes Vs Ages with status');
plt.show();
```

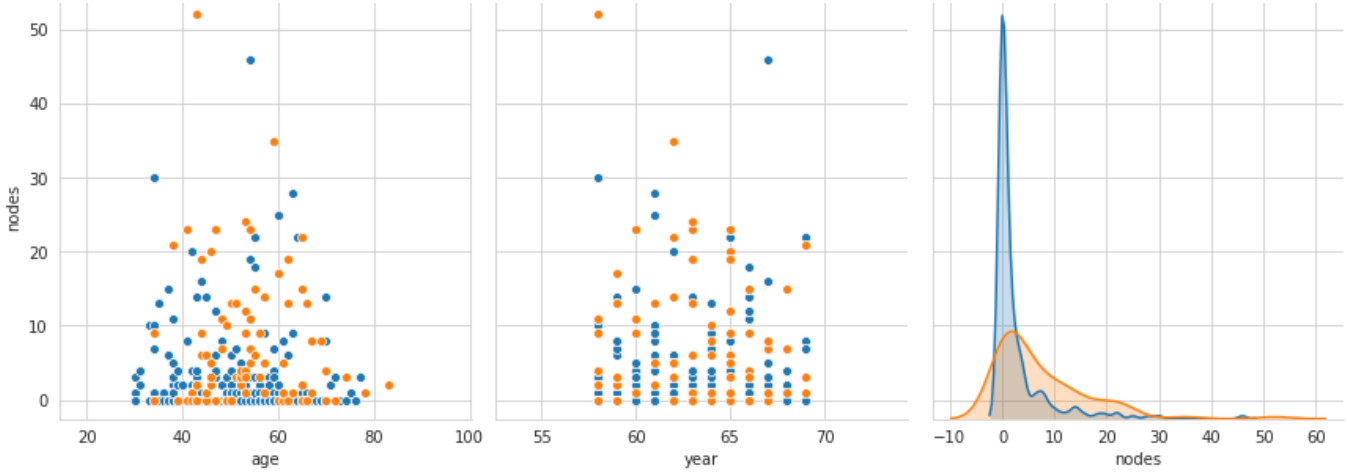



**PAIR PLOT**

In [ ]:

```
plt.close();
sns.set_style("whitegrid");
sns.pairplot(data, hue="status",size=4)
plt.show()
```



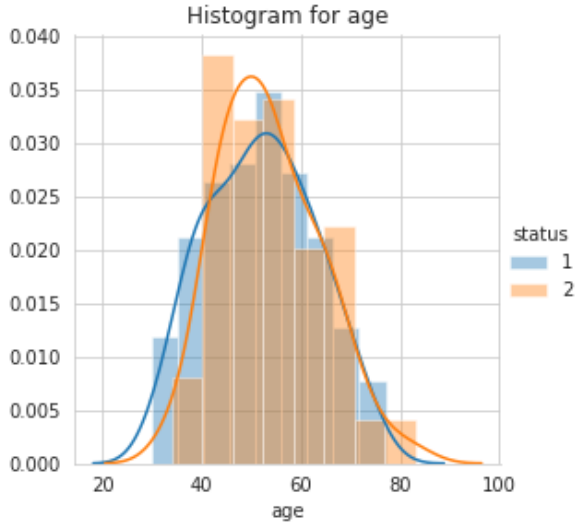



**observation:most of the data is overlapping**

**Histogram,pdf ,cdf**

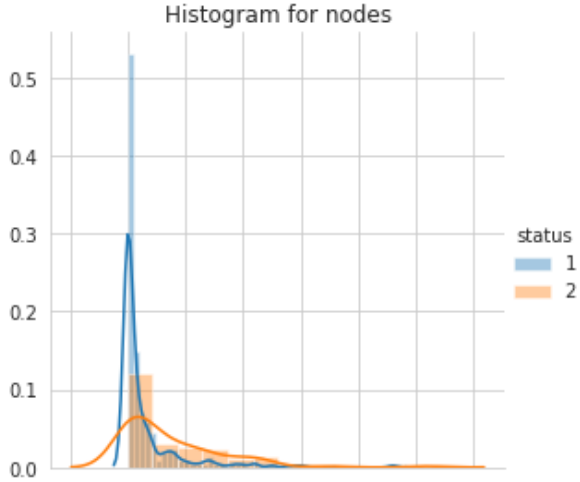
In [ ]:

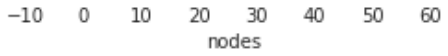
```
sns.FacetGrid(data, hue="status", size=4) \
   .map(sns.distplot, "age") \
   .add_legend();
plt.title('Histogram for age');
plt.show();
```




In [ ]:

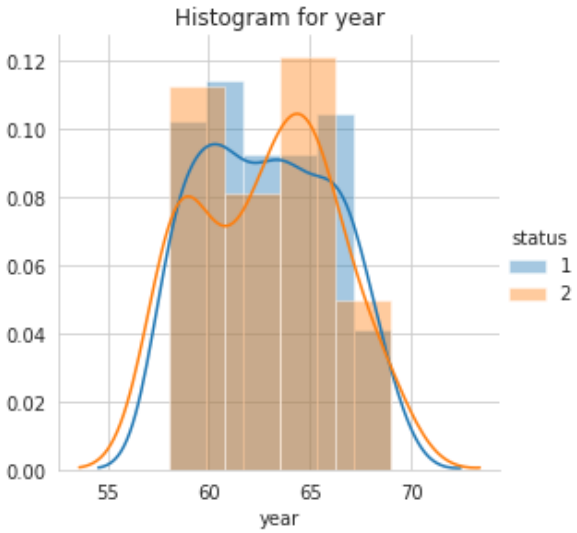
```
sns.FacetGrid(data, hue="status", size=4) \
   .map(sns.distplot, "nodes") \
   .add_legend();
plt.title('Histogram for nodes');
plt.show();
```






In [ ]:

```
sns.FacetGrid(data, hue="status", size=4) \
   .map(sns.distplot, "year") \
   .add_legend();
plt.title('Histogram for year');
plt.show();
```




**CDF**

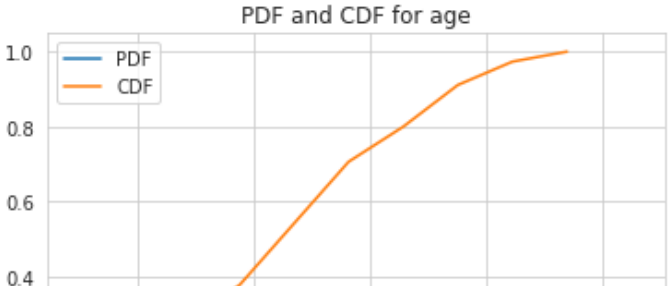
In [23]:

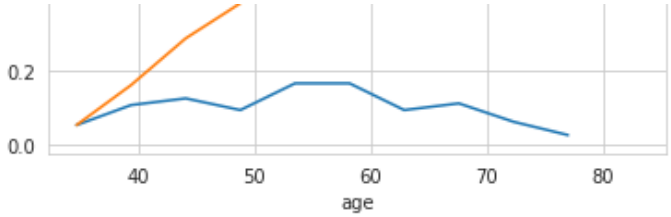
```
not_dead=data.loc[data["status"]==1]
dead=data.loc[data["status"]==2]
```

In [55]:

```
counts, bin_edges = np.histogram(not_dead["age"], bins=10,
                                 density = True)
pdf = counts/(sum(counts))
print(pdf);
print(bin_edges);
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:],pdf,label='PDF')
plt.plot(bin_edges[1:], cdf,label='CDF')
plt.legend()

counts, bin_edges = np.histogram(dead['age'], bins=1,
                                 density = True)
pdf = counts/(sum(counts))
plt.plot(bin_edges[1:],pdf);
plt.xlabel('age')
plt.title('PDF and CDF for age')
plt.show();
```

```
[0.05333333 0.10666667 0.12444444 0.09333333 0.16444444 0.16444444
 0.09333333 0.11111111 0.06222222 0.02666667]
[30.  34.7 39.4 44.1 48.8 53.5 58.2 62.9 67.6 72.3 77. ]
```

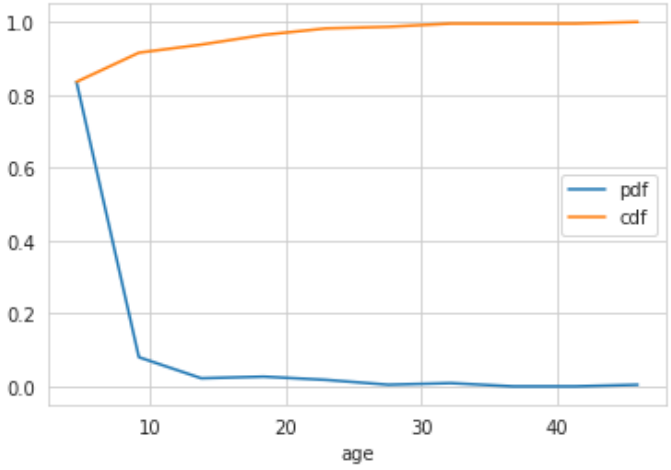





In [59]:

```
counts, bin_edges = np.histogram(not_dead['nodes'], bins=10, density = True)
pdf = counts/(sum(counts))
print(pdf);
print(bin_edges)
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:],pdf,label='pdf')
plt.plot(bin_edges[1:], cdf,label='cdf')
plt.legend()
plt.xlabel('age')
plt.show();
```

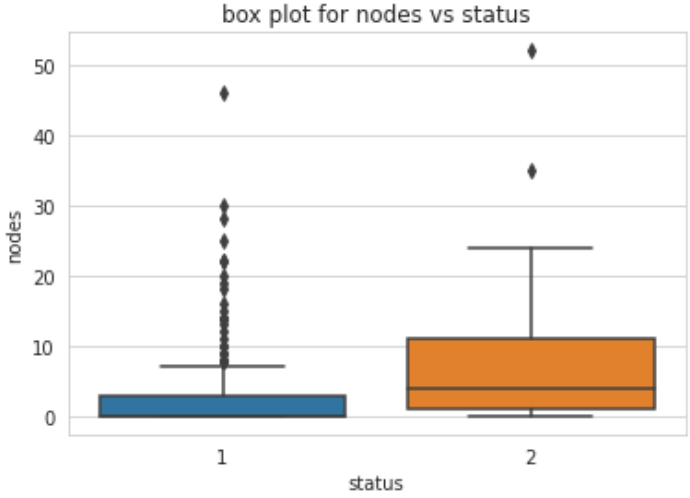
```
[0.83555556 0.08       0.02222222 0.02666667 0.01777778 0.00444444
 0.00888889 0.         0.         0.00444444]
[ 0.   4.6  9.2 13.8 18.4 23.  27.6 32.2 36.8 41.4 46. ]
```




**box plot**

In [35]:

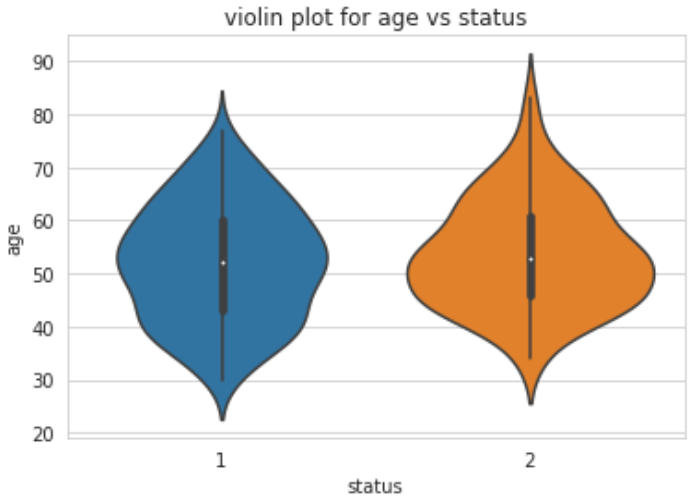
```
heberm=data
sns.boxplot(x='status',y='nodes',data=heberm)
plt.title('box plot for nodes vs status')
plt.show()
```




**violin plot**

In [36]:

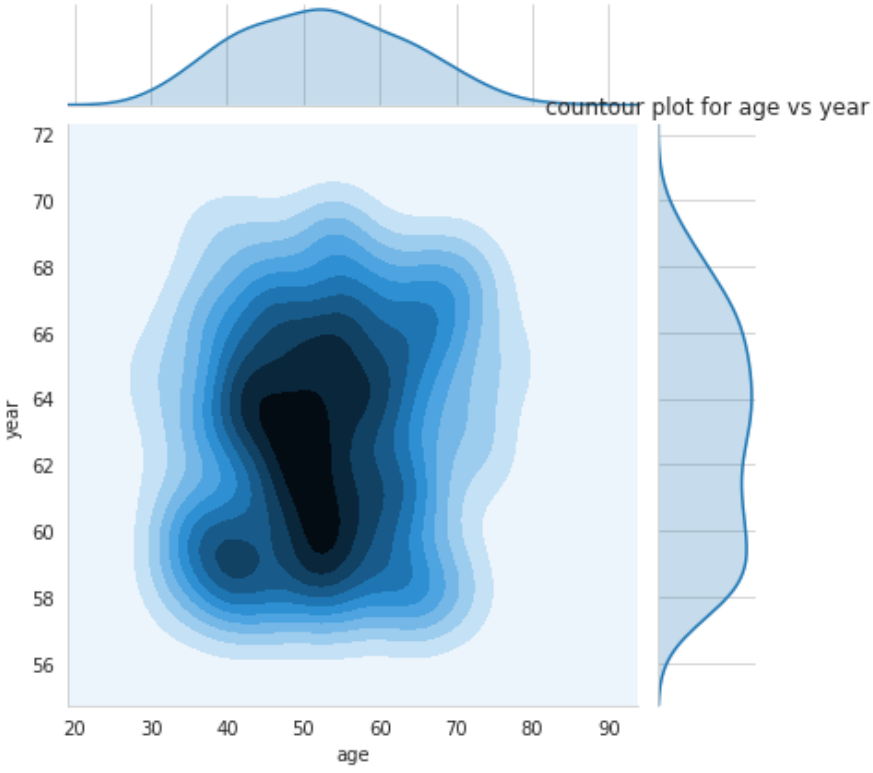
```
sns.violinplot(x="status", y="age", data=heberm, size=8)
plt.title('violin plot for age vs status')
plt.show()
```




**Countour plot**

In [37]:

```
sns.jointplot(x="age", y="year", data=heberm, kind="kde");
plt.title('countour plot for age vs year')
plt.show()
```




**obeservation:Countour plot shows that,between 1958-1964,plenty of people have gone through operation,between age of 50 to 60**